CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Short interval change in hepatitis C hypervariable region 1 in chronic infection. Are there treatment windows in the envelope?
Authors	Schmidt-Martin, Daniel
Publication date	2016
Original Citation	Schmidt-Martin, D. 2016. Short interval change in hepatitis C hypervariable region 1 in chronic infection. Are there treatment windows in the envelope? PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2016, Daniel Schmidt-Martin. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2024-03-29 13:03:23
Item downloaded from	https://hdl.handle.net/10468/4632

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Short interval change in Hepatitis C Hypervariable Region 1 in chronic infection

Are there treatment windows in the envelope?

Volumes: 1

Author: Daniel Schmidt-Martin MB BAO BCh MMedSci

A thesis submitted for consideration for PhD in Medicine

National University of Ireland, Cork

Department of Medicine, National University of Ireland, Cork, Ireland.

Submitted: March 2016

Head of Department: Professor Fergus Shanahan

Supervisors:

Dr Liam J Fanning

Dr Orla Crosbie

Dr Elizabeth Kenny Walsh

Table of Contents

Table of Contents i

Declaration vii

Summary viii

Acknowledgements x

Contributors xi

Chapter 1 – 1

Molecular Virology of Hepatitis C Virus 1

1.1 Background to Hepatitis C Virus 2

1.1.1 Introduction 2

1.1.2 Classification 2

1.1.3 Hepatitis C Genotypes 3

1.1.4 Clinical Relevance of HCV Genotypes 3

1.1.5 Geographical Distribution of Genotypes 4

1.1.6 Origins of HCV 4

1.1.7 Recombination 5

1.1.8 HCV Epidemiology 5

1.1.9 Mode of Transmission 6

1.1.10 Iatrogenic Hepatitis C Infection in Ireland 6

1.1.11 HCV in Egypt 7

1.2 Natural History of Hepatitis C Infection 7

1.3 Hepatitis C virus 9

1.3.1 Molecular Structure of HCV 9

1.3.2 5’ Untranslated region (UTR) 9

1.3.3 3’ Untranslated Region 9

1.3.4 HCV structural proteins 10

1.3.5 Core 10

1.3.6 Envelope Proteins 11

1.3.7 Hypervariable Regions 12

1.3.8 HVR1 13

1.3.9 HVR2 13

1.3.10 HVR3 13

1.3.11 P7 14

1.3.12 HCV Non Structural Proteins 15
1.3.13 NS2 15
1.3.14 NS2/3 16
1.3.15 NS3A 16
1.3.16 NS4A 18
1.3.17 NS4B 18
1.3.18 NS5A 19
1.3.19 NS5B 20
1.4 HCV Life Cycle 21
1.4.1 Cell entry 21
1.4.2 Translation and processing of HCV polyprotein 22
1.4.3 HCV replication 22
1.4.4 Virus assembly and release 23
1.5 HCV and the Immune System 24
1.5.1 Evasion of the Innate Immune Response 24
1.5.2 Evasion of the Adaptive Immune Response 25
1.6. Quasispecies 27
1.6.1 Background 27
1.6.2 Quasispecies bottlenecks 28
1.6.3 HVR1 Quasispecies 28
1.6.4 HVR1 Quasispecies in Acute HCV infection 28
1.6.5 HVR1 in Chronic HCV Infection 29
1.6.6 HVR1 Post Transplantation 30
1.6.7 HVR1 in Co infection with HIV 31
1.6.8 HVR1 Quasispecies on Treatment 31
1.6.9 HVR1 as an Immune Target 31
1.7 Treatment 33
1.7.1 Background 33
1.7.2 Treatment Individualisation 33
1.7.3 Viral Predictors of SVR using dual therapy with Pegylated Interferon and Ribavirin 34
1.7.4 Host Predictors of SVR using Dual Therapy with Pegylated Interferon and Ribavirin 37
1.8 Conclusion 40
Chapter 2 42
Methods 42

2.1 Methods 43
2.1.1 Aims 43
2.2 Study Design 45
2.2.1 Subject Recruitment 45
2.2.2 Consent Form 47
2.2.3 Patient Attendance 52
2.3 Methods in Chapter 4 53
2.3.2 RNA extraction and amplification 53
2.3.3 Reverse Transcription 53
2.3.4 Nested PCR of E1/E2 53
2.3.5 PCR amplicon of the correct size was confirmed using gel electrophoresis. 55
2.3.5.1 2% agarose gel preparation 55
2.3.5.2 Loading Samples and Running an Agarose Gel: 55
2.3.5.3 Recipe for TAE 55
2.3.6 Cloning 56
2.3.7 Preparing Agar plates 56
2.3.8 Amplification of Cloning Plasmid 57
2.3.9 Preparation of samples for sequencing 57
2.3.10 Sequence analysis 58
2.3.11 Calculation of nucleotide substitution rates 58
2.3.12 Statistical Analysis 59
2.4 Methods in Chapter 5 60
2.4.1 Taq Polymerase nested PCR 60
2.4.2 PCR product Purification 60
2.4.3 Sequence analysis 61
2.5 Methods in Chapter 6 62
2.6 Methods in Chapter 7 63
2.6.1 Ultradeep pyrosequencing data generation, handling and error correction. 63
2.6.2 1-step and k-step network construction. 63
2.6.3 Bioinformatics analyses. 64
2.7 Methods in Chapter 8 65
2.7.1 Separation of sample into Immunoglobulin enriched/ depleted fractions 65
2.7.2 RNA extraction and nested E1/E2 PCR 66
2.7.3 Partitioned analysis of Quasispecies. 66

2.8 Methods in Chapter 9 67
2.8.1 Interferon gamma inducible protein 10 (IP 10) assay 67
2.8.2 RT PCR 67
2.8.3 Core nested PCR 67
2.8.4 Interferon Sensitivity Determining Region (ISDR) PCR 68
2.8.5 IL28 Sequencing 70
2.9 Summary of Project Outputs 71
Chapter 3 72
Published Hypothesis Paper 72
Hepatitis C quasispecies adaptation in the setting of a variable fidelity polymerase 72
Chapter 4 81
Published Study 81
Intensive temporal mapping of Hepatitis C hypervariable region 1 quasispecies provides novel insights into HCV evolution in chronic infection. 81
Chapter 5 105
Does amplicon sequencing provide accurately reflect the underlying HVR1 quasispecies? 105
5.1. Background 106
5.2. Methods 106
5.3. Results 107
5.3.1 Subject B 107
5.3.2. Subject C 109
5.3.3 Subject D 111
5.3.4 Subject F 113
5.3.5. Subject K 115
5.3.6. Subject L 117
5.3.7. Subject M 119
5.3.8. Subject N 121
5.3.9. Subject Q 123
5.3.10. Subject T 125
5.3.11. Results Summary 127
5.4. Discussion 130
5.5. Conclusion 131
Chapter 6 132
Analysis of long term HVR1 sequence evolution 132

6.1 Background 133
6.2 Methods 133
6.3 Results 133
6.3.1 Subject C 135
6.3.2 Subject F 142
6.3.3 Subject H 149
6.3.4 Subject Q 156
6.3.5 Results Summary 163
6.3.5.1 Time order Phylogeny 163
6.3.5.2 HVR stasis 164
6.3.5.3 Number of Subpopulations 164
6.3.5.4 Nucleotide Substitution Rates 165
6.4.1. Discussion 169
6.4.2 Diversity, Complexity, and Divergence 169
6.4.3. Phylogenetics 170
6.4.4 Sequence Alignments and Subpopulations 170
6.4.5 Substitution Rates 171
6.4.6 E1 vs HVR1 Substitution Rates 172
6.5. Conclusion 174
Chapter 7 175
Published Study 175
Network analysis of the chronic Hepatitis c virome defines HVR1 evolutionary phenotypes in the context of humoral immune responses 175
Chapter 8 189
Analysis of IgG binding patterns and influence on HVR1 sequence change 189
8.1 Introduction 190
8.2 Methods 190
8.3 Results 191
8.3.1 Subjects with no HVR1 change 192
8.3.2 Subjects with HVR1 change 197
8.4 Discussion 200
8.5 Conclusion 201
Chapter 9 202

Predicting response to pegylated interferon and ribavirin using HVR1 quasispecies parameters in conjunction with known viral and host markers 202
9.1. Introduction 203
9.2 Methods 203
9.3. Results 204
9.3.1 Core and ISDR Sequences 204
9.3.2. Temporal Change in IP 10 207
9.3.3. Pre Treatment Prediction Data Summary 211
9.3.3. Genotype 1 patients 211
9.3.3. Genotype 3 patients 211
9.4. Discussion 212
9.4.1. Combination studies investigating prediction of treatment response. 213
9.5. Conclusion 215
Chapter 10 216
10.1 Conclusion 216
Appendix A 219
Genbank Accession Numbers 219
Appendix B 231
Retrospective data for remaining Subjects 231
Bibliography 333

Declaration

I certify that this thesis has not been previously submitted for a degree in this or any other University. This thesis is the result of my own investigations and any other assistance is acknowledged.

Signed _____ Daniel Schmidt-Martin

Summary

Hepatitis C Virus (HCV), an RNA virus, is one of the leading causes of cirrhosis worldwide and, remains the leading indication for orthoptic liver transplantation in the United States.

Dual treatment with pegylated interferon and ribavirin has until 2010 been the mainstay of treatment. The emergence of newer agents with direct activity against specific virus proteins has revolutionised HCV treatment but, the high cost of these medications are likely to prevent universal access, particularly in developing countries and, strategies to optimise response to cheaper combination treatments are required. The Irish Hepatitis C outcomes research network (ICORN) has proposed a target of 2025 for the complete eradication of Hepatitis C from Ireland.

HCV replicates in an error prone fashion resulting in mutant progeny known as quasispecies(QS), thought to form an important mechanism of host immune evasion in the establishment and maintenance of chronic infection, which develops in 50-80% of those acutely infected.

HCV has three hypervariable regions (sections of the virus genome that appear to tolerate higher substitution rates) and one of these, Hypervariable region 1 (HVR1) has been recognised as a major target of the adaptive immune response. HVR1 quasispecies complexity and diversity have been implicated as predictive of response to dual therapy. Little, however, is known about the natural history of these parameters in chronic infection.

We discuss evolutionary concepts and how they apply to quasispecies and hypothesise how viruses might select a setting appropriate mutation rate in order to optimise adaptation, advancing the theory of replicative homeostasis.

We prospectively study 23 patients with chronic HCV infections and, differing degrees of liver fibrosis fortnightly for a 16 week period prior to commencement of treatment. Using amplicon sequencing, cloning and next generation sequencing we explore the behaviour of HVR1 QS, establishing the utility of each technique in describing QS change.

We identify variable and unpredictable HVR1 change in our cloning data which precludes the use of these metrics in pre treatment prediction models. HVR1 change is far greater in non cirrhotic patients and the transition to cirrhosis appears to be associated with a change from positive to purifying selection. Using molecular clock techniques we illustrate differing substitution rates within HVR1 among cirrhotic and non cirrhotic patients.

We identify, by including an additional retrospective sample, that the patterns we describe are sustained over prolonged periods and further clarify the mode and tempo of HVR1 change by estimating the substitution rates.

Using next generation sequencing techniques we identify similar patterns of HCV change when compared with our cloning data. However, the sequence depth provided permits the description of time specific network of HVR1 clones, all connected by a single amino acid substitution to a central node.

By separating our samples into immunoglobulin bound and free fractions we describe the importance of host immune mediated change driving the changes seen in our pyrosequencing and cloning data.

Finally, using known viral and host molecular markers predictive of treatment response we explore unsuccessfully for models predictive of treatment response.

Acknowledgements

I would like to thank my supervisors Dr Liam, Fanning, Dr Orla Crosbie, and Dr Elizabeth Kenny-Walsh for their continual support, advice and understanding in bringing this project to completion. I would also like to thank Professor Fergus Shanahan for his insightful and measured advice, in addition to his infectious optimism and enthusiasm.

I would like to thank Liam and Brendan Palmer for their time and helpful, entertaining, and invaluable assistance in challenging me to continually raise the bar in the pursuit of academic achievement.

To my colleagues in the Department of Medicine and those from the Department of Pharmacology: Bernie, Jacquie, Kevin, Brendan, Isabelle, Mary Jane, Brendan, Peter, Trish, Peter, Grace, Mike, Philana, Aileen, Caitriona, John Stack, Beth and of course John Levis what can I say? Thanks for guiding me at the bench, advising me when I needed it and consoling me when it went wrong. Thanks for listening to me, entertaining me and teaching me. In particular to Kevin and Brendan thank you for the endless distraction of discussions about hurling.

I would like to thank my colleagues in the Department of Hepatology who assisted in identifying suitable candidates for the study and provided invaluable assistance in sample collection. To Susan Corbett, I will be forever indebted to you for your inestimable knowledge of the patients and their stories.

To the patients, without whom this project would not have been possible, I give my sincere thanks. I will forever remember the lengths patients went to in order to provide samples and will fondly remember how the ash cloud threatened to completely disrupt the project as a number of people became stranded abroad. I feel honestly humbled by the efforts made to provide samples and the life lessons learned from hearing and experiencing the life stories of each patient. To those for whom we were able to clear the virus I rejoice with you. To those for whom we did not succeed for you, I hope that this research can provide answers and pledge to fight for you and your health.

Finally, I must thank my family for unwavering support. To Mum and my two brothers and sister thank you for listening and supporting me at all times. To my wife Mary thank you, thank you, thank you: for your love, your patience, your encouragement, your understanding, your advice and for never doubting me. To my children Beth, Eoin, and Maebh: I love you and promise not to do a post doc.

Contributors

Daniel Schmidt-Martin drafted the study proposal and successfully applied for grant support for the project from Molecular Medicine Ireland. Completed the application for ethical approval. Recruited all patients and collected all samples. All experiments were performed by Daniel Schmidt – Martin with the exception of minor assistance in the IP 10 assay. Performed all data preparation, analysis and interpretation for all aspects of the project with the exception of the pyrosequencing where bioinformatic support was provided. Daniel Schmidt-Martin co-authored the paper investigating HVR1 networks using pyrosequencing and authored all other project outputs.

Liam Fanning assisted in the planning and scope of the project. He advised with regard to laboratory techniques and assisted in the drafting of all project outputs. Liam also provided support in the overall interpretation of results and contextualisation of the findings within the current understanding of hepatitis C virus evolution.

Orla Crosbie assisted in the drafting of the thesis and provided support for all clinical aspects of the project supervising patients as they were treated.

Elizabeth Kenny Walsh assisted in the drafting of the thesis and provided clinical support during the treatment of all subjects.

Susan Corbett assisted in identifying suitable candidates for the study and co-ordinated patient attendance for samples to be taken.

Brendan Palmer provided bioinformatic support for the analysis of the pyrosequencing outputs and co authored the paper describing the network analysis of short interval change in HVR1 sequences. Brendan also provided invaluable trouble shooting in the event of experimental challenges.

John Levis and Bernie Crowley provided assistance in technical aspects of RNA extraction and nested PCR.

Shake it Off

Regrets collect like old friends

Here to relive your darkest moments

I can see no way, I can see no way

And all of the ghouls come out to play

And every demon wants his pound of flesh

But I like to keep some things to myself

I like to keep my issues strong

It's always darkest before the dawn

And I've been a fool and I've been blind

I can never leave the past behind

I can see no way, I can see no way

I'm always dragging that horse around

And it's hard to dance with a devil on your back
So shake him off

I tried to dance with the devil on your back

And given half the chance would I take any of it back

It's a final mess but it's left me so empty

It's always darkest before the dawn

And I'm damned if I do and I'm damned if I don't

So here's to drinks in the dark at the end of my road

And I'm ready to suffer and I'm ready to hope

And it's hard to dance with a devil on your back

So shake him off **F. Welch**

Chapter 1 –

Molecular Virology of Hepatitis C Virus

Predicting response to pegylated interferon and ribavirin using Hepatitis C quasispecies

1.1 Background to Hepatitis C Virus

1.1.1 Introduction

Hepatitis C is a positive stranded RNA virus, the first member of the *Hepacvirus* genus which includes related viruses hosted by bats, primates, rodents, horses, and cows, all members of the *Flaviviridae* family of viruses. Initial observations describing patterns of infective hepatitis described two differing patterns, one with a short incubation period which was enterally transmitted and a second parenterally transmitted infection which had a longer incubation period(1). With the isolation and identification of two major causative agents, Hepatitis A an RNA virus within the *Picornavirus* family and Hepatitis B virus a DNA virus from the *Hepnaviridae* family it became clear that further infective agents were causing a proportion of the chronic infective hepatitis. Hepatitis C virus had as a result been previously included under the umbrella term of non A non B hepatitis (NANB) prior to the isolation and identification of the causative agent by Michael Houghton's group at Chiron group in 1989(2). With the development of robust serological testing it emerged that HCV was responsible for up to 90% of NANB hepatitis with the majority of the remainder accounted for by Hepatitis E and Hepatitis G infection(3, 4).

1.1.2 Classification

Hepatitis C was placed within the *Flaviviridae* family of viruses, which includes Flaviviruses, and Pestiviruses, due to overall similarities in the genomic structure and replication strategies. The *Flaviridae* family comprises up to 80 known viruses divided into three genera with antigenically distinct characteristics(5):

Flavivirus

Flaviviruses are tick borne RNA viruses and include Yellow Fever, Dengue, and Japanese Encephalitis Virus(5).

Pestivirus

This genus includes a number of RNA viruses that infect ruminant animals and include Bovine Viral Diarrhoea Virus 1 and 2, Classical Swine Fever Virus, and Border Disease Virus which affects sheep(6).

Hepacvirus

Hepatitis C virus was until recently the only virus in this genus, though GB B virus a positively stranded RNA virus of marked similarity to HCV has been recently included as a second member(7). There are

three GB viruses (A ,B ,C) and GB virus B is the one with most sequence identity (28%) to HCV(8). GB B virus is hepatotropic and can infect a number of monkey species, though the true host remains unknown(9). Gb-B virus has been proposed as a potential model for studying HCV and a chimeric HCV/GB-B construct has recently been patented(10). GB-C virus is lymphotropic and recent data has suggested that co infection with HIV may reduce the pathogenicity of HIV(11).

1.1.3 Hepatitis C Genotypes

Early studies of HCV genomic composition demonstrated marked heterogeneity in nucleotide sequences. This prompted the subdivision of the genus into 11 differing genotypes but due to variable techniques used to define these, the redefinition of HCV into a universally accepted classification has been undertaken(12). As a result HCV was divided into 6 genotypes differing from each other by 31-33% at the nucleotide level. Within each genotype there are subtypes which differ by 20-25% at a nucleotide level(13). Subsequent study has resulted in the description 7 genotypes and 67 subtypes of HCV in the most recent classification(14).

1.1.4 Clinical Relevance of HCV Genotypes

HCV genotypes have important clinical implications with differing genotypes associated with differing patterns of HCV disease progression and treatment response. In addition to requiring treatment for twice as long as genotypes 2 and 3, genotypes 1 and 4 respond to combination pegylated interferon plus ribavirin in 30-42% of cases whereas genotypes 2 and 3 achieve sustained response to treatment in 70-80%(15) (Fig 1.). Recent efforts do describe the evolution of HCV genotypes suggest that the most interferon resistant genotypes (1 and 4) may represent the most recent evolutionary change in HCV(16). Additionally genotype 3 is strongly associated with abnormal lipid metabolism leading to hepatic steatosis, a major risk factor for developing fibrotic liver disease(17). Retrospective analysis of patients who became infected by multiple blood transfusion or through intra venous drug use (IVDU) suggest that infection with multiple genotypes may occur in up to 19% and 3-9% of cases, respectively(18, 19).

The first significant advance in HCV treatment since the development of pegylated interferon and ribavirin occurred with the release in 2011 of telaprevir and boceprevir. These directly acting HCV protease inhibitors demonstrated significant efficacy against genotype 1 and have demonstrated potential efficacy in genotype 2 but no efficacy in genotype 3(20). These first generation protease inhibitors were beset by significant side effects and associated with significant adverse events in patients with advanced liver disease, including a number of mortalities and have subsequently been replaced with next generation protease inhibitors and polymerase inhibitors with improved pan genotypic response rates and more favourable side effect profiles, though the improvements in

genotype 3 patients have been modest, particularly in cirrhotic patients(Fig. 1). Combination treatments have, for the first time, resulted in interferon free regimens for most genotypes(21-24).

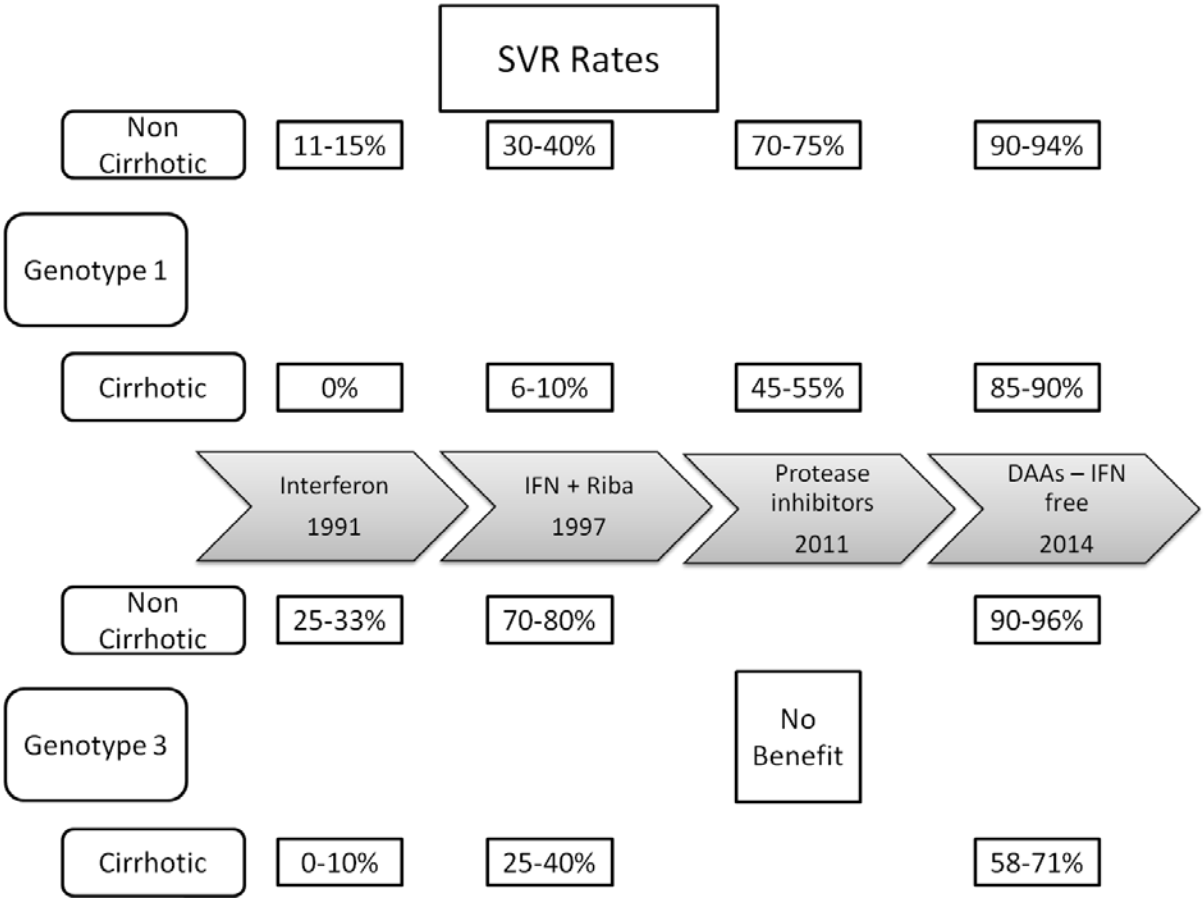


Fig 1. Advances in treatments and response rates to HCV treatment in genotype1 and 3 patients with and without cirrhosis.

1.1.5 Geographical Distribution of Genotypes

With increasing population migration the geographical distribution of HCV genotypes is continually changing, most especially in destination countries such as the United States and in Western Europe. Genotype 1 accounts for a high proportion of infections in the United States and Central Africa where genotype 4 is also high in prevalence. Western Africa has a high prevalence of genotype 2 and genotypes 3 and 6 are most commonly seen in South and Eastern Asia(13).

1.1.6 Origins of HCV

The date of introduction of HCV to the human population remains unclear with estimates ranging from several hundred to several thousand years. Evidence from the number of genotype 2 and 6 variants

in Central/Western Sub-Saharan Africa and South East Asia respectively point to long term endemic infection(25, 26).

1.1.7 Recombination

Although most HCV can be classified into the 7 genotypes described in the most recent classification of HCV, a number of replication viable variants made up of recombined sequences from different genotypes have been described. First identified in a 2k/1b recombinant in Russia, recombination challenged original theories that HCV evolution occurred mainly by the accumulation of point mutations and provides insight into how the new genotypes are likely to emerge as it potentially facilitates the exploration of remote sequence space(27). Quantification of intra-host recombination events is challenging, but evidence of intergenotypic recombination in affecting other genotypes Peru, Ireland, The Philippines and Uzbekistan suggest that recombination is more common than previously assumed(27-31).

1.1.8 HCV Epidemiology

Estimating the global burden of hepatitis C has proven challenging but, it is estimated to chronically infect 120-170 million worldwide, roughly equivalent to 3% worldwide prevalence with a significant proportion unaware of their infection(32).

Precisely quantifying hepatitis C prevalence is challenging for a number of reasons.

1. Many estimates of prevalence are based on seropositivity within at risk populations and large scale population studies are rare, particularly in developing countries.
2. In countries where recreational intravenous drug use has become the predominant mode of transmission, limited interaction with healthcare professionals in this cohort may serve to underestimate prevalence. The use of outreach programmes and screening of active users at drug centres appears to offer the prospect of improving estimates of HCV prevalence among these patients (33-35).
3. Few patients develop clinically obvious acute infection and as a result few patients present to health care providers at the time of initial infection.

HCV is now the leading indication for liver transplantation in the US. Extrapolation of the healthcare costs associated with HCV projects healthcare related costs arising from the complications of HCV infection will rise from $6.5 billion in 2014 to a peak of $9.1 billion in 2024 with the current costs of eradication in the United States estimated at $80.1 billion(32, 36, 37).

1.1.9 Mode of Transmission

Hepatitis C is transmitted through bodily fluids, predominantly blood or blood products and sexual transmission among long term monogamous partners is unusual. Prior to the identification of HCV, blood transfusion of blood and blood products resulted in up to 40,000 iatrogenic infections per annum in the United States alone. With the advent of serological testing and the availability of screening tests there has been a change in the pattern of transmission in many developed countries with the rise in intravenous drug use supplanting iatrogenic infection. Nonetheless, the failure to implement (or the haphazard implementation of) adequate blood product screening and the widespread reuse of hypodermic needles in healthcare settings has meant that iatrogenic remains a significant mode of transmission in many developing countries. IVDU has replaced iatrogenic infection as the main cause of infection in Europe, North America, and Australia(32).

One of the challenges noted in explaining HCV spread is identifying how the virus was spread in the absence of a known animal host and before the advent of modern healthcare practices in the twentieth century. Indeed, studies of HCV genetics in certain areas of Sub Saharan Africa suggest endemic HCV infection may have been present in the population for several centuries. Ritual tattooing and certain religious practices such as circumcision have been implicated in both historical and ongoing transmission of HCV, as indeed, has insect bite born transmission. As HCV is unable to infect or replicate in arthropods, this would require the rapid transfer of virions carried in arthropod mouthparts(38).

1.1.10 Iatrogenic Hepatitis C Infection in Ireland

In Ireland, an estimated 1000 Irish women were exposed to HCV contaminated anti-D immunoglobulin between 1977 and 1978. The exposure first came to light following the introduction of blood product screening in 1991 which demonstrated an abnormally high proportion of Rhesus negative donors with anti HCV antibodies. Retrospective review of donor medical histories indicated that almost all had received anti D immunoglobulin between 1977 and 1978. (39, 40) The source of the outbreak was subsequently isolated to a plasma donor who had been diagnosed with infectious hepatitis. The inoculate was subsequently isolated and is regarded as a unique instance worldwide of a known time of infection of a specific HCV genome and it has been proposed as suitable for gaining insight into HCV evolution and natural history. Follow up clinical data on this cohort indicated limited progression of liver disease with as few as 2% developing cirrhosis at 17 years post exposure(41). Coinciding with the discovery of the initial Anti-D immunoglobulin associated outbreak, a second exposure dating to the period between 1991 and 1994 affecting 44 women was identified(42).

A second large cohort of 2,867 women with iatrogenic exposure to HCV via contaminated Anti-D immunoglobulin between 1977 and 1978 in the former East Germany has been followed up with 25 year clinical follow up data available. This group demonstrated comparable rates of RNA positivity (46%) and a low rate of cirrhosis development (0.5%), only one hepatocellular carcinoma and a HCV related mortality rate of 0.5%(43).

1.1.11 HCV in Egypt

Egypt has the world's highest seropositivity prevalence at 22% with rates as high as 55% in certain population demographics(44). This epidemic has been attributed to the reuse of hypodermic needles during a national anti schistosomal treatment programme undertaken between 1960 and 1980(45). It is estimated that as a result, there will be more than 117,000 deaths from hepatocellular carcinoma attributable to HCV in Egypt between 2008 and 2028(46).

1.2 Natural History of Hepatitis C Infection

1.2.1 Acute Hepatitis C

Acute hepatitis C, although frequently associated with non specific symptoms such as fatigue, nausea, abdominal pain, anorexia, pruritis or myalgia, most patients remain asymptomatic and cases of fulminant liver failure due to acute HCV infection are rare(47, 48). HCV RNA becomes detectable between 7 and 21 days post infection though liver blood test abnormalities often post date this by 8 to 12 weeks by which time HCV RNA levels have risen rapidly(49, 50). The emergence of anti HCV antibodies occurs in most patients between 32-46 days post transmission, though in immunocompromised patients this may be delayed up to 48 weeks(51, 52). Acute HCV infection may be cleared spontaneously in a minority of cases with pre menopausal women, patients with acute infection characterised by a clinical hepatitis with jaundice and, patients with favourable nucleotide polymorphisms adjacent to the interleukin 28B gene most likely to clear the virus without treatment(53-55).

1.2.2 Chronic Hepatitis C

According to studies of seropositive populations, acute HCV infection progresses to chronic infection in between 64-78% of cases(32, 54, 56). Chronic infection, once established, often follows a relatively indolent course with patients seldom presenting to healthcare professionals. HCV itself is not thought to be pathogenic to hepatocytes rather it is the immune response to the virus that causes the complications associated with HCV infection, though there is a growing body of evidence suggesting that virus proteins directly modulate a number of cellular pathways affecting lipid metabolism, insulin

sensitivity, and apoptosis pathways(57-59). Although predominantly hepatotropic, extra hepatic replication of HCV has been demonstrated in both lymphocytes and within the central nervous system(60, 61).

Ongoing immune mediated hepatic inflammation which is characterised by marginal and often intermittent elevations in liver enzymes are thought to result in gradual fibrotic change in the portal tracts which, can lead to portal hypertension and cirrhosis. Although liver function tests can estimate current inflammation, imaging modalities such as ultrasound and Fibroscan and biopsy evaluation of liver tissue are often required to definitively quantify both inflammation and the development of fibrotic change. Cirrhosis develops in 20% of cases after 20 years and, with the onset of cirrhosis, there is an associated increased risk of developing hepatocellular carcinoma which appears to be of the order of 1% per annum(62). The presence of other co morbidities such as obesity, other pathologies such as additional hepatic pathology or lifestyle risk factors such as excessive alcohol consumption may hasten the advance of end organ damage in HCV(32). Once established, cirrhosis is associated with a severe clinical complication rate (ascites, variceal bleeding or encephalopathy) and a mortality rate of 5.6-8.3% and 2-4% per annum respectively(63-65). Successful eradication of HCV prior to the onset of cirrhosis reduces patient mortality rates to that of the general population(66). The advent of decompensation or a potentially curative hepatocellular carcinoma should prompt consideration for liver transplantation but, patient and graft survival rates for patients post transplant are inferior to those transplanted for all other indications(67).

1.3 Hepatitis C virus

1.3.1 Molecular Structure of HCV

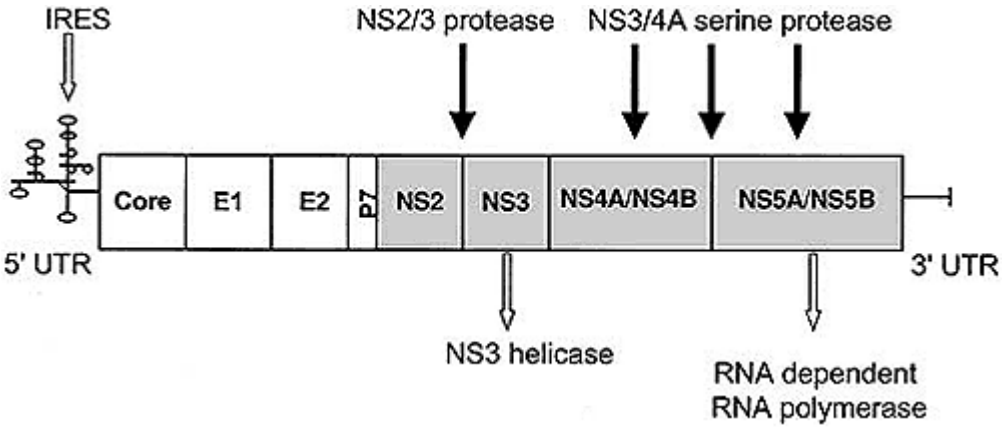


Fig. 2. Molecular structure of HCV virus.

1.3.2 5' Untranslated region (UTR)

HCV is a 9,600 base pair single stranded RNA virus. The 5' terminal of the HCV genome is composed of a 341 nucleotide nontranslated region which is the most highly conserved region of the genome (90.1% sequence identity)(68). Although nontranslated, this portion of the genome forms a 4 domain complex secondary structure which is essential for its function(69). Together with the first 24-40 nucleotides from the N-terminal of the core protein, domains II, III and IV form an internal ribosome entry site (IRES) which is capable of initiating cap independent translation of the HCV genome. The IRES binds host 40S ribosomal subunits and guides this host translational machinery to the methionine AUG initiation codon(70). Domain I (nucleotides 4-20) is not essential for replication or translation, but may have a modulatory function(71). Finally, recent studies of HCV sequences has shown an association between variation at nucleotide 243 located in domain III and numerous nucleotide positions within NS2 and NS3 non-structural proteins that are required for HCV translation and replication. Cell culture assays using subgenomic particles with these variations have indicated a possible role for NS2 in modulating the rate of replication(72).

1.3.3 3' Untranslated Region

The 3' UTR is a tripartite sequence of variable length made up of a highly variable 30-50 nucleotide section, a poly-U/UC tract of variable length (20-200 nucleotides) and completed by a 98 nucleotide highly conserved χ-tail(73, 74). Similar to the 5'UTR, the 3' UTR sequences form secondary structures which are important in the efficient replication of HCV(75). HCV is unable to replicate both in cell culture and in vivo in the absence of the poly-U/UC tract or the χ-tail and, replication has been shown to be significantly reduced in the absence of the variable region(76, 77). The χ-tail forms 3 stem loops,

the second of which has been shown to interact with stem loops in NS5B, most likely, in the initiation of minus strand transcription(78, 79). Modification of the χ-tail has been shown to reduce replication efficiency(80). Finally, disruption of the 3'UTR appears to significantly reduce IRES dependent polyprotein translation(81).

1.3.4 HCV structural proteins

1.3.5 Core

Core is a 21 kDa, 173-179 amino acid α helical protein which demonstrates marked intergenotypic genomic conservation. The first protein in the HCV open reading frame, it is initially cleaved by a host signal peptidase creating a 23 kDa 191 amino acid immature form and achieves its mature form following post translational C-terminal processing. Conserved hydrophilic bases at the N-terminal are responsible for homo-oligmerisation, forming the nucleocapsid and enclosing and binding genomic RNA which are the primary structural functions of the protein. During translation a sequence motif located between core and E1 results in the attachment and subsequent translocation of E1 across the endoplasmic reticulum (ER) membrane with concomitant cleavage of the core-E1 junction and release of core which itself, becomes associated with lipid droplets via domain 2(82, 83). Core association with lipid droplets has been demonstrated at numerous intracellular sites including membranous webs, the surface of lipid droplets and on the membranes of the endoplasmic reticulum. Variable post translational modification of core has been associated with differing patterns of localisation with palmitoylation, required for interaction with ER membrane(84). Core has been strongly implicated in disruption of lipid metabolism, both in cell culture models and mouse models, and it has been proposed that core is at least in part responsible for the steatosis seen in hepatitis C through a reduction in lipid droplet turnover(85, 86). The association of core with lipid droplets is required for the production of infectious HCV particles and, it has been shown that this association is required for virus assembly (87-89).

In addition to modulation of lipid metabolism, HCV core protein has a number of other putative regulatory functions. A number of cell culture and animal studies have demonstrated an upregulation of apoptotic and apoptotic-like pathways through activation of Fas, mcl, and tumor necrosis factor(59, 90-93). Indeed, a role for core protein in the genesis of hepatocellular cancer has also been suggested(94). In vitro studies have demonstrated that core proteins can increase reactive oxygen species and result in mitochondrial stress/dysfunction(95). Cell culture studies have also demonstrated a possible link between core protein and insulin resistance often seen in patients with hepatitis C through the disruption of insulin receptor substrate 1, a known target of insulin(96, 97).

Finally, core protein has shown the potential for manipulation of host immune response through the disruption of interferon signalling and the alteration of Kupffer cell function(98, 99).

Further truncated forms of core protein produced as a result of alternative reading frames have been isolated but the relevance of these remain somewhat elusive(100). Recently, cell culture studies have demonstrated that an alternative reading frame present in the core genome can result in the translation of so called "minicore" proteins ranging from 8-16 kDa and lacking the N-terminal amino acids. The precise role and functions of these proteins remain to be clarified(101).

1.3.6 Envelope Proteins

HCV has two envelope proteins designated E1 [35kDa] and E2 [70 kDa] which are cleaved from the HCV polyprotein by host peptidases(102). Our understanding of envelope protein structure and function had been limited due to the restrictions of the *in vitro* pseudo particle model which produced replication competent, but not infectious particles. The advent of a cell culture model has contributed greatly to the elucidation of envelope protein structure and function(103). E1 and E2 are type 1 transmembrane proteins with large N-terminal ectodomains and short intraluminal C- terminals(102). E1 and E2 undergo post translational glycosylation at 6 and 11 sites respectively at the N-terminal(104). This most likely occurs at the golgi apparatus and this is thought to be required for envelope protein folding and virus particle assembly, and appear to have functions in both CD81 binding during cell entry and, in disrupting the effectiveness of host anti E2 neutralising antibody response(103, 105). Amino acid similarity across genotypes for the envelope proteins is 68%(106).

Until recently E1 and E2 were thought to form a functional unit comprised of non-covalently bonded heterodimers which were thought to form the viral envelope(107). However, following the publication of a proposed secondary structure of E2 it appears that these heterodimers may be present intracellularly but that the envelope is made up of more complex E1:E2 oligomers which are formed by covalent disulphide bridges(103, 108). This is in agreement with previous studies suggesting that E2 is the fusion protein, similar to the class II fusion proteins seen in other *Flavi* and *Alphaviruses*, required for HCV cell entry(108). However, analysis of the secondary structure of E2 did not elucidate typical conformational features of class I, II, or III fusion proteins. The precise role for E1 at that time was unclear. Although it was recognised that anti E1 antibodies prevented HCV cell entry, there was no evidence for direct interaction between E1 and host cell receptors(109, 110). It had been proposed that E1 functions as a lattice structure, ensuring the correct conformational folding of E2 and, facilitating the generation of the oligomers required for envelope building. Coimmunoprecipitation

studies have recently identified binding of E1 but not E2 to apolipoproteins which are thought to bind host LDL receptors(111). The secondary structure of E1 also does not however, conform to class I, II, or II fusion proteins suggesting that E1 may be a novel form of fusion protein (112). Most recently E1 homotrimers have been described on the virus surface and, it has been proposed that these are formed using E2 as a co factor and, facilitate the internalisation of the virus with disruption of these homotrimers associated with a loss of virus infectivity(113).

E2 has been shown to bind directly with CD81 and to interact with a number of other receptors which have important roles in cell entry. Evidence that the envelope proteins are required for cell entry has meant that this has been a target for the development of HCV vaccines though results have been disappointing(114). The development of anti envelope antibodies has been shown *in vivo* both in acute and chronic infection. Antibody development in acute infection has been associated with a lower likelihood of spontaneous clearance and antibodies are almost ubiquitous in chronic infection(115). A number of studies have identified potential broadly neutralising antibodies targeting E2 against conserved regions of E2 in animal models (114, 116, 117). Recently, a panel of broadly neutralising antibodies has been successfully used to prevent HCV infection, and abrogate infection in chimeric mouse models potentially identifying a novel strategy for eradicating infection, as it is dependent on continual infection of new hepatocytes(118). Ineffective antibody responses are thought to be related to the heterogeneity of envelope sequences, which is thought to facilitate immune evasion. Other factors that are thought to minimise immunogenicity of the E1:E2 glycoproteins are: the closely covalently bound conformation that form which minimises protein exposure; the multiple glycans that protect their exterior; and the association between virion surface glycoproteins and host lipids(103, 119, 120). E2 contains 3 regions designated hypervariable regions 1,2 & 3, which have been proposed as antibody binding sites(121). E2 binding with CD81 receptors on natural killer cells *in vitro* indicates that E2 may have an additional role in modulation of host defences by altering interferon signalling(122).

1.3.7 Hypervariable Regions

In two separate papers in February 1991, two areas of marked amino acid heterogeneity were described within, what was then, the putative envelope encoding region of the HCV genome(123, 124). These sections of the genome have been designated the hypervariable regions (HVR) 1 and 2 and are located between amino acids 384-411 near the junction between E1 and E2, and amino acids

476-482 respectively. Recently, a third hypervariable region has been described (HVR3) located between amino acids 431-466. All three HVRs share a number of physico chemical properties with marked variability tolerated but also site-specific conservation of certain amino acids which are thought to ensure correct functional folding of the protein product with a number of basic residues required for cell entry(121, 125).

1.3.8 HVR1

HVR1 is a 27 amino acid section of the N-terminal of the E2 protein which forms a tail like structure proximal to domain 1, in the recently proposed tertiary structure of the HCV E2 protein(108, 123). HVR1 is the most variable but, also demonstrates amino acid conservation at particular sites and an overall positive charge which it has been proposed points to a role in cell targeting or binding(121, 126). HVR1 has been proposed as a B and T cell epitope. HVR1 appears to have three microdomains with amino acid residues at positions 14, 15 and 25-27 essential for binding of the virus to scavenger receptor class B, type1 receptor. This efficiency of binding to these receptors appears to be modulated by amino acids 1-13. The third microdomain encompassing amino acid residues 16-24 is not required for cell entry and is an epitope for neutralising antibodies(127).

1.3.9 HVR2

HVR2 comprises a 7-11 amino acid segment of E2 and it demonstrates 39-93% sequence identity depending on the genotype(128). Serum sampling studies have shown that HVR2 undergoes less sequence divergence than HVR1 which has led to speculation that it is not a target for immune surveillance(126). HVR2 has been shown to overlap with the binding site of CD81, a tetraspanin receptor required for HCV cell entry, though in vitro studies have demonstrated that binding of CD81 can occur in the absence of HVR2, albeit at reduced efficiency(128). On treatment studies of sequence changes in HVR2 have demonstrated it to be under selective pressure but, one study of substitution patterns in HVR2 quasispecies during treatment failed to demonstrate any significant difference between responders and non responders(129-131). This has led to speculation that HVR2 may protrude from E2 and provide a protective shield to the true CD81 binding site which is exposed through conformational change in the period immediately prior to cellular binding(128). Interestingly HVR2 provides two glycosylation sites which have also been implicated in shielding HCV from host immune response(119).

1.3.10 HVR3

HVR3 is a 17-36 amino acid portion of E2 which is located between HVR1 and HVR2 with 3 subdomains, which has been shown to be under strong host selective pressure and, is thought to function in the process of HCV binding(132, 133).

Finally, two further hypervariable regions within E2, but confined to genotype 3a, have recently been identified. HVR495 and HVR575 are, respectively, 7 and 9 amino acids in length, appear to be under selective pressure, and are flanked by conserved hydrophobic residues. The functional relevance of these regions is unknown, but they correspond to potentially important folding sites for the E2 glycoprotein (108, 134).

1.3.11 P7

This 63 amino acid hydrophobic polypeptide with two helical transmembrane domains, and a conserved basic cytosolic loop orientated toward the lumen of the endoplasmic reticulum is essential for *in vivo* production of infectious particles, but not required for replication(135-137). Analogous proteins have been described in BVDV(138). Similar to viroporins, it has recently been shown to form hexamers, and to function as an ion channel ,which may protect pH sensitive virus particles, most likely glycoproteins, during virion assembly(139, 140). Alternate processing of HCV polypeptide results in the production of E2:p7 fusion proteins which, though not required for production of infectious progeny, may be incorporated into virus particles(141).

1.3.12 HCV Non Structural Proteins

The non-structural proteins are coded for the 3' or C-terminal two thirds of the HCV genome. They are designated non-structural because they do not form part of the circulating HCV particle but, a number of them perform structural functions during the translation and replication of HCV. The HCV non-structural proteins also contain a number of co factors and two enzymes required for replication(142). HCV does not have a non-structural protein 1 as once classified as a Flavivirus it was found not to have an analogous non-structural protein(143).

1.3.13 NS2

NS2 is a 23 kDa hydrophobic transmembrane protein, with three transmembrane segments and localises to the endoplasmic reticulum, into which it is inserted through its N-terminal(144). The timing and mechanism of translocation and insertion into the membrane remain to be fully elucidated, though it has been proposed to occur both co and post translationally, once cleavage of NS3 has occurred(144, 145).

NS2 has been implicated in the hyperphosphorylation of NS5A which is required for NS5A activity though, it may be that this accreditation is erroneous and that it is NS3 that is responsible for this function(146, 147). More recently, it has been shown that host casein kinase II (CK2) may be responsible for this hyperphosphorylation(148). Interestingly NS2 is also hyperphosphorylated by CK2 and is rapidly degraded by the proteosome thereafter(149). Whether this action is coincidental or either non-structural protein is required for these actions remains to be fully elucidated.

NS2 has shown potential for regulation of host immune responses by inhibiting a number of promoters which are associated with pro inflammatory cytokines suggesting a role in the control of host immune response(150). NS2 has been shown to prevent apoptosis by preventing localisation of CIDE-B (cell death inducing DFF45-like effector) to the mitochondrion, where it functions as a pro-apoptotic protein(151).

Finally, NS2 has been shown to contain a site for the development of intergenotypic chimeras. These genotype switch over joints occur within NS2 following the first transmembrane domain suggesting that the NS2 N-terminal interaction with the structural proteins is required for viable replication(152). This intriguing finding has been further studied with NS2 co localisation with E2 close to lipid droplets on confocal microscopy, suggesting a role for NS2 in HCV virus assembly. Confocal microscopy and co immunoprecipitation demonstrate that NS2 interacts with E1, E2, P7, NS3, and NS5A and suggest that NS2 may uniquely provide the foundation for virus assembly at the endoplasmic reticulum(153, 154).

1.3.14 NS2/3

NS2/3, a highly hydrophobic protein extending from amino acids 810-1206, is the first non-structural protein to be translated(155). Cleavage from the structural protein p7 is performed by host signal peptidases. NS2/3 is a novel cysteine protease which functions as an autoprotease in subsequently cleaving NS2 from NS3 between amino acids 1027 and 1028(156-158). NS2/3 appears to form dimers which may be required for protease activity, with each protein providing either one or two of the amino acids required for the active cleavage site(158, 159). Following cleavage the C-terminal residues remain at the active site and it has been suggested that this may serve to inactivate the virus(160). The hydrophobic N terminal of NS2 is not required for autoprotease function and the function of this domain appears to relate to actions of NS2 is virus assembly following autoproteolytic cleavage(144). The full NS3 protease domain must be present for NS2/3 processing, but this occurs independent of the serine protease activity of NS3(156).

NS2/3 conformation plays a crucial role in autoprotease activity and this can be affected by NS4A derived peptides which, by altering the conformation of the NS3 N-terminal, disrupt the positioning of the cleavage site(161). Zinc is an essential component of NS2/3 protease activity but it seems that rather than being a zinc dependent protease, that zinc is required for maintenance of the required structure(160). NS2/3 cleavage is required for viral replication in vivo but, the successful replication of subgenomic particles coding for NS3-3'UTR suggests that replication may occur in the absence of NS2(162, 163). HCV NS2/3 demonstrates sequence alignment similarity with Bovine Viral Diarrhoea Virus (BVDV) NS2/3. In BVDV, replication rates has been shown to correlate with cleaved NS3 levels whereas infectivity requires non cleaved NS2/3 suggesting a potential role for regulating replication and infectivity rates, though similar findings have yet to be demonstrated in HCV(160, 164, 165).

1.3.15 NS3A

The non-structural protein 3 (NS3)is a 70kDa multifunctional protein containing a serine protease at its C-terminal which makes up 2/3 of the 5' end of the protein and an RNA helicase of the DExH family at the 3' end(166). Both enzymes may act independently of the other but there is also evidence that either may modulate the activity of the other(167). Additionally, the both the protease and NS4A have been shown to regulate the activity of the helicase(168, 169).

The protease is a serine protease and member of the trypsin/chymotrypsin protease superfamily(170). Zn 2+ stabilises the structure of the protease and is essential for enzymatic function. The Zn2+ is itself located within a construct containing three cysteine molecules and a water molecule(171).

The activity of the NS3 protease requires a catalytic triad of conserved amino acids and an oxyanion hole. NS4A contributes to this by localising substrate and catalytic triad(172). NS3 is responsible for the cleavage of NS3/4a, NS4a/4b, NS4b/5a, and NS5a/5b to release the respective non structural proteins(173).

The NS3 protease cleaves both viral and host proteins and has been implicated in viral mechanisms for evading innate immune responses. NS3 cleaves both TRIF and MAVS both of which are important in the normal activation of the RIG-I pathway, which in turn upregulates host interferon stimulating genes and endogenous interferon activity, which has been shown to be important in initiation of HCV infection (174, 175).

Several NS3 protease inhibitors have been developed with the implication that inhibition of NS3 would both interfere with established infection by disrupting virus replication and additionally prevent/reduce intra host spread through the associated reactivation of innate immune responses(176). Two protease inhibitors (boceprevir and telaprevir) have been licensed for use in chronic hepatitis C infection and were associated with improved efficacy in genotype 1 infections but, have subsequently been withdrawn due to the emergence of newer agents with fewer side effects and improved treatment responses(177, 178). Interestingly, many of these inhibitors have encountered resistant mutant variants which are postulated to exist within the quasispecies at time of initial inhibitor exposure(179).

The 3’ third of the NS3 codes for a helicase which is a member of the DExH subfamily of DEAD RNA helices, which is itself part of subfamily 2, one of the two main families of helicases, as divided by amino acid similarity(180). It can act on RNA,RNA/DNA and dsDNA substrates and acts in a 3’to5’ direction(181). It is made up of three domains and the amino acid Trp-501 located in domain 3 is thought to anchor the protein to the substrate and predispose the directional mode of action of the protein(182). Helicases are thought to unwind the helical secondary structures formed by genetic material, thus facilitating translation and replication. It is thought to unwind helices in a 1-3 base pair kinetic step using NTPs and dNTPs as a source of energy. The precise mechanism for this remains unclear but, it has been proposed that while anchored by domain 3 that domains 1 and 2 move base pair by base pair along the genome with the hydrolysis of ATP providing the energy required for unwinding in what has been dubbed the inchworm mechanism. The separating of the DNA strands

appears to require the Beta hairpin which extends from domain 2, as the absence of this structure abolishes DNA unwinding in a spring loaded fashion. It is not clear whether the NS3 helicase operates as a monomer, dimer or oligomer as these conformations have all been demonstrated in vivo(183). Numerous helicases can cooperatively act by binding at different sites owing to their unidirectional mode of action. NS3 has an optimal activity at pH 6.5 and a binding site of 7-8 nucleotides(184). As mentioned previously, NS3 protease is involved in modulating the activity of NS3 helicase. The precise mechanism is not clear but as NS3 protease binds DNA with greater affinity than NS3 helicase, a role binding unwound DNA has been postulated(169).

1.3.16 NS4A

NS4A, a 54 amino acid protein, the shortest non structural protein has a number of functions in viral replication, host immune evasion and, virus assembly. It is required as a co factor in the NS3/NS4A enzyme complex which is essential for the cleavage of the non structural protein junctions at NS3/4A and NS4B/5A and optimises cleavage at NS4A/4B and NS5A/5B(185). A hydrophobic N terminal stabilises the NS3-4A complex to cellular membranes. The middle portion acts as a cofactor ensuring optimal folding of the NS3 serine protease domain(186). NS4A is also required for hyperphosphorylation of NS5A an essential process in viral replication(187).

The NS3-4A complex exhibits an ability to disrupt the innate immune response by cleaving the cellular messenger IPS-1, a membrane associated protein which forms part of the RIG-I mechanism of innate anti viral prevention of HCV infection(175, 188). Additionally, the NS3-4A complex has been implicated in virus assembly, possibly through the association between the acidic C terminal and host membranes at the endoplasmic reticulum(189).

1.3.17 NS4B

NS4B is a protein of molecular weight 27 kDa which is made up of 261 relatively conserved amino acids with a hydrophobic predominance(190). NS4B is formed by the cleavage by NS3-4A serine protease of NS 4A from a NS4A-NS5A complex and, the subsequent cleavage between NS4B and NS5A in what are the penultimate and final steps in the cleavage of the HCV non-structural proteins(191, 192). This is thought to occur at specific membrane sites in order to facilitate NS4Bs inherent transmembrane phenotype(193). NS4B contains four transmembrane regions and an N terminal which is thought to be initially cytosolic with the capacity for intraluminal translocation and a C terminus which is cytosolic(190, 194). NS4B localises to membranes both in the endoplasmic reticulum and, in dot like membrane aggregates seen in the cytoplasm of infected cells known as membranous

webs(MW)(195, 196). The translocation of the N terminal seems to occur as a result of oligomerisation, which results in the transformation of the second of the two N terminal α helices into a transmembrane segment(197, 198). Oligomerisation is required for MW formation and, it may that oligomerisation is required in order to achieve the required concave conformation of the MW(199). These conformational changes have also been shown to be influenced by alterations in the relative abundance of other HCV non-structural proteins, suggesting that the activity of NS4B may be in turn modulated by these(197, 200). The C terminus is characterised by two α helices, one of which, (α2) anchors the protein to the membrane(201). Interestingly, the C terminus is characterised by a high degree of amino acid conservation and has recently been demonstrated as an important factor in facilitating interactions between NS4B and itself in the formation of functional replication complexes(202). NS4B hydrolyses ATP and NTP an enzymatic function which may serve to provide energy autonomy to virus replication(203). NS4B has also been proposed to function in the assembly of virion particles, to bind RNA during replication, the disruption of innate immune activation and, in mitigating the anti viral activity of interferon α(204-207). Furthermore protein interaction networks have suggested a role for NS4B in inducing oxidative stress as a result of NS4B induced endoplasmic reticulum stress. Finally, network analysis of protein-protein interactions have also implicated NS4B in the development of steatosis, insulin resistance, liver fibrosis, and tumour development in HCV patients, though the mechanisms have yet to be elucidated(199, 208).

1.3.18 NS5A

NS5A is a 447 amino acid proline rich phosphoprotein which exists *in vivo* in two forms: p56 and p58 depending on the degree of phosphorylation the polyprotein has undergone(157). Phosphorylation appears to be mediated by NS4A but also appears to involve a number of host kinases, including casein kinase I α and Polo like kinase I. The precise role for the differing degrees of phosphorylation remains to be elucidated but, the NS5A appears to have important roles in a number of areas within the host cell including the cytosol and adjacent to the endoplasmic reticulum. NS5A is made up of three domains with domains I and II forming part of the replication complex and is involved in virus replication and domain III involved in virus assembly. Domain I includes both a Zinc binding site and a disulphide bond that bind RNA directly and, both of which are necessary for replication. The absence of domains I and II prevents virus replication and amino acid substitutions may have significant effects of the replicative capacity of the virus(209). Phosphorylation of domain III is required for virus assembly and this section co localizes with core proteins on lipid droplets and appears to regulate the

transition and encapsulation of new genomes from the endoplasmic reticulum to form new virus particles(209).

NS5A has been identified as a potential drug target and the first direct NS5A inhibitor Daclastavir has completed clinical trials and demonstrated high degrees of efficacy though a number of resistant mutants have already been described(210). A section of NS5A within genotype 1b has been associated in a number of studies on populations in the Far East with pre treatment prediction of response to treatment with dual therapy combining pegylated interferon and ribavirin. Named the Interferon Sensitivity Determining Region (ISDR), patients with wild type respond less frequently to treatment, while those with 2 or more amino acid substitutions at this site have significantly superior response rates(211).

1.3.19 NS5B

NS5B is a 591 amino acid 86 kDa protein which is cleaved by the NS3 serine protease(212). The N terminal 530 amino acid portion forms a classic fingers, palm and thumb subdomain motif that is seen universally in RNA dependent RNA polymerases. The RNA template binds directly to the groove between the fingers and thumb which leads directly to the active polymerisation site(213). There is significant pangenotypic conservation of NS5B which has lead to the development of a number of highly efficacious NS5B inhibitors which have revolutionised HCV treatment(214). NS5B requires both magnesium and manganese as co factors and is capable of polymerisation *in vivo*. The C terminal tail is composed of a 21 amino acid segment which binds to membranes and is not required for polymerisation in vitro but, is essential for replication in cells(213).

1.4 HCV Life Cycle

1.4.1 Cell entry

HCV transits the bloodstream bound to low density lipoprotein (LDL), very low density lipoprotein [VLDL], bound to immunoglobulin, and in free form. HCV cell entry is by means of a number of receptors. The envelope proteins E1 and E2 form a heterodimer functional unit, which was thought to interact and attach directly to the virus receptors though recent studies have suggested that the virus is bound to Apolipoprotein E which interacts with the hepatocyte LDL receptor(215). Both E1 and E2 are required for HCV receptor binding(216).

LDL receptor (LDLR) is a potential candidate as the virus transits bound to LDL but the role remains undefined. It may be that LDLR and glycosaminoglycans (GAGS) binding brings the virus in contact with the receptors required for internalisation(217). The C-type lectin receptors DC-SIGN (dendritic cell-specific intercellular adhesion molecule-3-grabbing integrin) and L-SIGN (liver cell-specific intercellular adhesion molecule-3-grabbing integrin) have also been implicated in cell entry but, their roles remain unclear(218, 219). CD 81, a tetraspanin, almost ubiquitously expressed, and scavenger receptor class B type I (SR-BI), a membrane bound lipoprotein receptor, bind E2 but, although required for cell entry, they are unable to transfer the virus across the membrane either alone or in combination (220-222).

Claudin-1 and occludin, tight junction proteins which form contacts between the apical poles of lateral cell membranes are also required for HCV cell entry. HCV is not thought to directly interact with the tight junction proteins, rather, following receptor binding, the virion receptor complex is thought to migrate to tight junctions, where internalisation occurs, though more recent data illustrating minimal association between Claudin-1 and CD81 at tight junctions and particle tracking studies that do not show migration of HCV particles to tight junctions suggest that tight junctions may not be required for cell entry(223-226). Epidermal growth factor receptor (EGFR) activation (by means yet to be elucidated) appears to stimulate the association of claudin and CD81 which form a co-receptor complex involved in virus cell entry(227, 228). Cell entry is by clathrin mediated endocytosis which, is pH dependent and the virion is incorporated into an endosome and delivered to the endoplasmic reticulum (229-231).

It is thought CD81 binding may induce conformational changes in the HCV virion, possibly to the E1:E2 envelope structure which primes the virus for the low pH environment required for internalisation(232). Infected cells down regulate claudin and occludin expression, which reduces membrane polarisation and is thought to prevent superinfection (233). Furthermore, HCV infection alters the localisation of tight junction proteins from plasma membranes to lateral membranes, thus

potentially facilitating virus transfer to neighbouring hepatocytes and compartmentalisation of infection within the liver(234, 235). Additionally, claudin 6 and 9 act as co-receptors in cell entry(236).

1.4.2 Translation and processing of HCV polyprotein

Following internalisation, fusion between the virion glycoprotein envelope and the cellular membranes is thought to occur releasing the uncoated genome into the cytosol(237). HCV genome contains an internal ribosome entry site (IRES) which is comprised of domains II, III, and IV of the highly conserved 5' UTR together with nucleotides 24-40 of the core gene. This IRES initiates cap independent translation by binding with the 40S ribosomal unit and the initiation factor eukaryotic translation initiation factor 3 (eIF3) (70, 238). Meanwhile, HCV eIF3 can also utilise eIF4F to initiate cap dependent translation(239). Furthermore, NS5A can upregulate cap dependent translation mechanisms(240). The 5' UTR also binds endogenous miR-122, an endogenous micro RNA which increases translation and replication(241). Finally, it has recently been shown that the 3' UTR can also enhance IRES activity, which may select for the translation of intact genomes(240). The translation of the single open reading frame produces a polyprotein which is both co- and post- translationally cleaved by viral and host proteases(101, 143).

1.4.3 HCV replication

HCV replication occurs at membranous aggregates called membranous webs that are thought to be derived from the endoplasmic reticulum and are mediated by the actions of NS3/4a, NS4b, NS5a and NS5b(195). These membranous webs are made up mostly of double membrane vesicles which are induced by NS5a(242). Replication is meditated by a combination of NS3 NTPase/helicase activity and the NS5B RNA dependent RNA polymerase (RDRP). Hyperphosphorylation of NS5A is thought to control the activity of the RDRP(148). The intracellular association of HCV with lipid droplets and the finding that replication in cell culture is disrupted by a change from saturated and monounsaturated fats to poly unsaturated fats has suggested a role for lipids in HCV replication(243). This has been reinforced by the recent finding that pharmacological manipulation of the lipid milieu toward HDL was associated with a reduction in HCV viral loads, though this was transient(244).

A complementary negative stranded genome is generated by the RDRP from which the positive stranded genome is subsequently produced. Newly generated positive strands are then either, translated, replicated or packaged into new viruses for release. The fate of the negative strand remains unclear(237). Host micro RNA 122 (miR-122) has been identified as an important in the stabilisation of viral RNA and prevents degradation by the viral 5' exonuclease Xrn 1 during replication by recruiting Argonaute 2(245). HCV induced alteration in host lipid metabolism results in the accumulation of intracellular lipids in the membranous web as lipid droplets which may have a role in viral replication

via interactions with NS5a(246, 247). Recently, SEC14L2, a single host cDNA that increases vitamin E modulated protection of viral particles from lipid peroxidation has been identified as the rate limiting factor which prevented the replication of HCV isolates other than the JFH-1 isolate in hepatoma cell lines(248). Finally, a number of further host proteins have been implicated in virus replication including phosphatidyl-inositol-4-kinase-III which is necessary for the formation of membranous webs and vesicle associated membrane protein associated protein A and B (VAP-A and VAP-B) which associate with cholesterol in the membranous webs(249, 250).

1.4.4 Virus assembly and release

Little is known of the precise mechanism by which HCV assembly and release is achieved. Once synthesized, core proteins homodimerize and are trafficked to lipid droplets where they accumulate and lead to transfer of lipid droplets to the peri-nuclear area (251-253). These core proteins are subsequently retrieved from the lipid droplets for virus assembly and budding in a process modulated by NS2 and p7(254, 255). NS2 seems to form a membrane bound matrix in order to facilitate the interaction of numerous structural and non structural proteins required for virion assembly (153, 154). E1 and E2 the envelope proteins form heterodimers and are maintained at the endoplasmic reticulum prior to NS2 mediated transfer for virion assembly, with p7 also required for capsid assembly and envelopment(256-258). A phosphorylated form of the C terminal of NS5a also plays a central role in virus assembly possibly by associating lipid droplet bound core proteins with NS5a and p7(259-261). Once enveloped, the virus exploits the very low density lipid (VLDL) assembly pathway and becomes associated with apolipoproteins(262, 263). It is thought that release is mediated by budding into the endoplasmic reticulum secretory pathway, though the direct cell to cell transfer mediated by intercellular tight junction proteins has also been proposed(217, 235).

1.5 HCV and the Immune System

1.5.1 Evasion of the Innate Immune Response

The discovery that certain patients display ability to clear HCV infection without sero converting prompted the investigation of innate immune anti viral mechanisms. HCV is sensed by all three main classes of pattern recognition receptors in the innate immune system:

1. Retinoic acid inducible gene-1 (RIG-I)

Spontaneous HCV clearance without seroconversion is mediated by activation of retinoic acid inducible gene-I (RIG-I) and is strongly associated with the host IL 28B (a host interleukin) genotype(264). RIG-I stimulates a signalling cascade that ultimately upregulates endogenous interferon signalling, which inhibits viral replication through the production of up to 300 interferon stimulated genes which induce an antiviral state in the host hepatocyte(265). RIG-I is a cytosolic RNA helicase which is activated early after infection by the poly U/UC at the 3' UTR of the HCV genome (264, 266-268). Interestingly, in patients where virus persists despite upregulation of interferon signalling, further augmentation with pegylated interferon is associated with high rates of treatment failure(269). Activation of RIG-I promotes oligomerisation and the activated complex is transferred to the mitochondrial associated endoplasmic reticulum membrane (MAM) where it interacts with mitochondrial antiviral signalling protein (MAVS)(270). This results in downstream activation of inflammatory molecules including nuclear factor kappa B (NFκB) and interferon regulatory factor 3 (IRF3)(188). HCV counteracts RIG I signalling by NS3/4A mediated cleavage of the mitochondrial antiviral signalling molecule (MAVS) and TIR-domain-containing adapter-inducing interferon-β (TRIF), both essential downstream components of RIG I signalling(265, 271).

2. Toll like Receptors (TLRs)

Toll like receptors may recognise either viral nucleic acid or protein(272, 273). Activation of TLRs reduces viral replication via TIR domain containing adapter-inducing IFN-β induced activation of IRF3 and NFκB. The HCV NS3/4A protease cleaves TRIF and downregulates TLR mediated innate immunity(174, 274, 275).

3. Nod-like Receptors (NLRs)

Nod-like receptors may sense HCV, though the precise pathogen associated molecular pattern responsible for their activation is unknown. NLR activation produces the pro-inflammatory cytokines IL-1β and IL-18(276-278).

The activation of the innate immune response although important in spontaneous clearance, is also an important step in priming and maturation of the adaptive immune response(265) .

1.5.2 Evasion of the Adaptive Immune Response

Adaptive immune response to HCV infection comprise both humoral (antibody mediated cell) and T cell potentiation of host anti viral interferon γ mediated mechanisms of viral clearance. Adaptive immune responses first become detectable 6-8 weeks after initial infection(279).

1.5.3 Humoral antibody mediated immune responses

HCV stimulates a variety of antibody responses to multiple viral epitopes but, these are predominantly non neutralising. The early emergence of a neutralising antibody response is associated with viral clearance and HLA restriction appears to play a significant role in identifying patients where spontaneous antibody mediated clearance is likely(280). The ability of HCV to stimulate a predominantly non neutralising antibody response may facilitate the virus in allowing time to exploit defects in the host antibody repertoire by the generation of escape variants, highlighting the importance of HCV genome malleability and its inherent quasispecies nature(265). Notably, hypogammaglobulinaemic patients may also clear HCV suggesting that antibody mediated clearances is not the sole mechanism of HCV clearance(281). HVR1 appears to be one of the main targets of anti HCV antibodies, though the protracted persistence of antibody bound HVR1 sequences highlights both the potential for crossreactive binding and the challenges the immune system faces in effectively clearing HCV(282-284).

1.5.4 T cell responses in acute and chronic HCV infection

Spontaneous clearance of acute HCV infection is associated with a robust CD4+ and CD8+ T cell response(285). Among Chimpanzee populations where HCV clearance is high, the depletion of either CD8+ or CD8+ T cell populations facilitates the persistence of chronic HCV infection until such time as the T cell populations recover(286, 287).

The mechanisms of viral evasion of T cell responses in both acute and chronic HCV infection are incompletely understood. Viral escape by means of genome mutation has been illustrated as an effective mechanism of escape at HLA epitope sites and is associated both with persistence of infection and spontaneous viral clearance. These immune escape mutants often require clustered amino acid substitutions at the epitope binding site which is likely to explain why individuals with the same HLA polymorphisms do not all spontaneously clear HCV infection(288, 289).

Further proposed mechanisms of viral evasion of T cell responses include dysfunction of CD8+ T cell response with resultant impaired interferon γ release and reduced T cell proliferation(290). Weak CD4+ T cell responses which are required to potentiate CD8+ T cell effector function have also been identified. Finally, it has been suggested that intrahepatic T cell regulatory cells may interfere with the proliferation of CD8+ cells, limiting the overall effectiveness of host T cell response(265).

1.6. Quasispecies

1.6.1 Background

Fundamental to the understanding of HCV evolution and immune evasion, which is a major factor in the maintenance of chronic infection, is the concept of quasispecies. The HCV RNA dependent RNA polymerase lacks a proof reading function and, as a result, the likelihood of replication of the genome into identical progeny is low. The HCV RNA dependent RNA polymerase has an estimated mutation rate of 10^{-4}/nucleotide site/year which is roughly equivalent to one mutation per replication cycle(291, 292). As many HCV virions are produced daily ($1x10^{12}$), this results in the generation of a highly heterogenous swarm or cloud of virions with differing genotypic and phenotypic characteristics. These collections of mutant virions are known as quasispecies. This term had originally been coined in relation to a theory proposing to explain the origins of self replicating organisms and, how selection and adaptation could be incorporated into early biological systems(293).

Fundamental to this theory was the generation of a constant proportion of mutant progeny in order to explore for beneficial mutations which might confer evolutionary advantage. Thus, the quasispecies generated would appear organised around a dominant sequence or master sequence which would represent the genome best adapted to the state of the quasispecies spectrum at a given time. Quasispecies theory differs from classical population genetics however in that the characteristics and behaviour of the quasispecies is seen as an ensemble property which incorporates both cooperative and competitive effects and, ultimately as a result the process of selection occurs at a population wide as opposed to an individual genome level. It has however subsequently been adopted and adapted by virologists and used to both explore and explain features of virus evolution, adaptation and selection(294, 295). With time the stepwise accumulation of mutant progeny allows the virus to test potential mutants also referred to as the sequence space and the maturation of quaspecies has been proposed as a state of equilibrium when constituent parts of the quasispecies become maintained in relative prevalence in proportion to the fitness conferred by each mutation.

While the generation of mutant progeny can result in virus evolution and adaptation, excessive mutation can result in an inability to maintain a master sequence. The point at which mutation rate exceeds the capacity for the quasispecies to maintain itself is referred to as the error threshold. Once a virus exceeds this threshold it enters into a process known as error catastrophe where the ever increasing number of progeny containing defective mutations results in a collapse in the quasispecies structure(295). This theory in the form of the lethal mutagenesis hypothesis, where viruses are induced into exceeding the error threshold, forms one of the major strategies for anti viral treatments(296). Perhaps the most important clinical implication of viral quasispecies is the capacity

of the virus to adapt to both new immune pressure and also, to maintain a reservoir of viral genomes that might contain mutants resistant to anti viral drugs(297).

1.6.2 Quasispecies bottlenecks

The transmission of viruses to new hosts often results in new infections founded by a random collection of virus genomes which may or may not contain master sequence genomes. This process, known as a bottleneck, has the potential to significantly reduce virus fitness if the sequences contain multiple deleterious mutations. Studies using serial bottleneck events in both viruses and phages have demonstrated an associated decrease in viral fitness validating the concept of Mullers ratchet which states that, in asexual replication, that the progeny of sequences with deleterious mutations are likely to contain the same mutation(298, 299). Conversely, bottleneck events if sufficiently infrequent and if containing sufficiently large numbers of viruses potentially facilitate an increase in viral fitness, by permitting the rise of sequences that had been the subject of interference from the dominant sequences within the preceding milieu(300).

Bottleneck events occur frequently in HCV with host to host transmission frequently characterised by the transfer of large numbers of viruses thus abrogating the risk of fitness loss. However, the requirement to infect individual host hepatocytes is a further potential bottleneck which can significantly affect the quasispecies profile of the infecting virus(301).

1.6.3 HVR1 Quasispecies

As described previously, HVR1 is a 27 amino acid section of E2 one of two envelope proteins on the virus surface (124). HVR1 is one of three regions in the HCV genome which demonstrates more marked sequence heterogeneity and dynamic change over the course of chronic infection when compared with the remainder of the genome, though certain structural motifs and the positioning of positively charged amino acid residues are conserved pointing to a role for HVR1 in cell attachment(121). Many studies have investigated quasispecies dynamics for HVR1 as it is postulated as a target for host immune response. Neutralising antibodies targeting HVR1 have been demonstrated and it is thought that HVR1 sequence change is driven by envelope targeting adaptive immune responses and that the malleability of HVR1 is important in facilitating the maintenance of chronic infection(302).

1.6.4 HVR1 Quasispecies in Acute HCV infection

The study of acute HCV infection has been limited by the indolent nature of the early stages of infection. This means that most diagnoses are made once chronic infection has been established. Exceptions to this rule are the limited number of patients who develop an acute hepatitis

characterised by jaundice, and among populations engaging in high risk behaviours (intra venous drug use (IVDU)) who were prospectively screened for acute infection. Transmission among IVDUs appears to be associated with an inoculum containing multiple HCV quasispecies(303). This contrasts with the recent outbreak of HCV among HIV infected men who have sex with men (MSM) where a lower diversity transmission has been demonstrated and a higher rate of spontaneous clearance reported suggesting that the diversity of the inoculum may be important in determining the likely outcome of acute infection(304). Acute infection is characterised by rapid changes in HVR1 quasispecies that appears not to be related to immune pressure and rather reflects the adaptation of the virus to optimise fitness (305). A proportion of patients will undergo rapid spontaneous clearance of HCV without the development of a humoral immune response. Seroconversion results in a reduction in HVR1 QS diversity suggestive of immune mediated clearance by neutralising antibodies(306, 307). In a landmark study Farci *et al.* identified an association between early QS HVR1 evolution and the development of chronic infection. Equally, a reduction in complexity and diversity was associated with increased likelihood of spontaneous clearance. Interestingly, Farci references the emergence of poly or multi phyletic trees as a predictor of viral persistence(308). This term is likely to correspond with the more recently described phenomenon of viral subpopulations(309). Acute infection is also associated with rapid early sequence change towards consensus sequences for each genotype, which is likely reflective of convergent change towards genotype specific fitness optima(310). The transition to chronic infection has been proposed to correspond with the failure of both innate and adaptive immune mediated clearance and, appears to be associated with an acceleration in the change in the HVR1 quasispecies, as the virus evades neutralising antibodies and CD4+ and CD8+ T lymphocyte responses. This is characterised by higher rates of non synonymous substitutions within the HVR1 region (310-313). Higher genetic complexity and diversity in HVR1 appears to be associated with the development of chronic infection, and it is postulated that HVR1 could act as a decoy, designed to induce immune response while protecting areas of the viral genome less tolerant of conformational change(313).

1.6.5 HVR1 in Chronic HCV Infection

With the transition to chronic HCV infection, HVR1 evolution is patient specific with variable patterns of change in HVR1 diversity, complexity, divergence and evolution. These fluctuate over time with no definitive pattern(314). Studies using paired liver biopsies suggest that serum quasispecies may lag changes in the liver by some weeks(315). In patients where treatment has been unsuccessful, early changes in HVR1 appear to be driven by selective pressures and result in significant diversification of

the quasispecies which suggest the exploitation of niche defects in the host immune system. The long term evolution of HVR1 appears to conform to two patterns: divergence from consensus sequence at the site of HLA epitopes and convergence towards consensus elsewhere in HVR1(316), which is likely to reflect the adaptation of the virus back to global fitness optima. These patterns of change result in the emergence over time of groups of HCV genomes that have been called both lineages and subpopulations in the literature (309).

The pattern of HVR1 change in chronic infection appears to have clinical implications, with those patients where HVR1 changes rapidly more likely to progress to advanced liver fibrosis and cirrhosis. These patients are also more likely to demonstrate ongoing active hepatitis in the form of raised liver enzymes, which implies active immune mediated hepatocyte damage(317). These episodic elevations of liver function tests also appear to correlate with transient spikes in the HCV viral load(318). With progressive fibrosis and cirrhosis HVR1 divergence and evolution appears to slow and there may be a reduction in HVR1 quasispecies complexity and diversity though studies in this area have included limited numbers of patients(319, 320). The transition to chronic infection appears to result in a reduction in HVR1 change as a result of the exhaustion of T cell and antibody mediated host immune responses which have been proposed to result from an element of original antigenic sin(321). Although many studies have investigated HVR1 change in chronic infection, little is known about the mode and tempo of HVR1 change over short time intervals.

1.6.6 HVR1 Post Transplantation

Recurrence of chronic HCV infection post orthoptic liver transplantation is almost universal and offers the prospect of evaluating quasispecies in an immune suppressed state. Patients transplanted for complications of hepatitis C have a poorer long term survival when compared with those transplanted for other indications(67). HCV recurrence in the transplanted organ may be asymptomatic in up to 50% of cases but may also be associated with rapid development of cirrhosis and a condition called fibrosing cholestatic hepatitis in up to 20% of cases. HVR1 complexity appears to be lower post transplant and greater degrees of post transplant HVR1 complexity appear to be associated with a more indolent post transplant course as does rapid change in the HVR1 quasispecies(322-324). Greater HVR1 diversity also appears to predict less progression in the liver allograft(325) though greater HVR1 divergence is associated with fibrosing cholestatic hepatitis(326). Finally, immunosuppression has been proposed to reduce immune pressure on the post transplant HVR1 quasispecies(327).

1.6.7 HVR1 in Co infection with HIV

A number of studies have suggested that immunosuppression is associated with reduced HCV HVR1 quasispecies complexity. Patients co infected with Human Immunodeficiency Virus (HIV) have historically represented a special group due to lower response rates to treatment with interferon and ribavirin but the emergence of new directly acting anti viral medications has resulted in equal response rates compared to patients not infected with HIV. Evidence for differing HVR1 quasispecies evolution/complexity, diversity and divergence has and remains both conflicting and controversial. A number of early studies using short single chain polymorphism analysis had suggested lower degrees of HVR1 complexity in patients co infected with HIV(328, 329). Subsequent studies evaluating the effect of immune reconstitution following the initiation of highly active antiretroviral therapy suggested that this resulted in increased HVR1 diversity and divergence with greater evidence of selection in the form of higher dn/ds ratios(330-332). These findings have partially conflicted with another study where no such association was found with sequence divergence(333).

1.6.8 HVR1 Quasispecies on Treatment

Rapid change in HVR1 quasispecies early in the treatment of HCV with dual therapy and, most particularly, the homogenisation of the quasispecies milieu has been shown to be predictive of treatment response(334). Furthermore, study of genotype 1 patients for whom dual therapy did not achieve sustained response has shown that those patients where the initial treatment resulted in a change in the HVR1 quasispecies had a greater chance of SVR following repeat treatment. This study was performed on patients receiving 24 weeks of non pegylated interferon and has important implications for our understanding of HCV quasispecies. Firstly, the lack of change among the cohort who failed treatment on two occasions suggested a persistent resistance to treatment which may have reflected either an inability of the host immune system to mount a neutralising immune response following potentiation of interferon signalling or, inherent resistance of the viral quasispecies to treatment. Secondly, the sub optimal duration of treatment suggests that the patients who cleared the virus following the second course of treatment may have cleared the virus if treatment had been extended to 48 or even 72 weeks which have subsequently been shown to further increase SVR rates. Finally, and most importantly, it suggested that the changes induced by the first course of treatment did not select treatment resistant mutants that would render future treatment futile(335).

1.6.9 HVR1 as an Immune Target

HVR1 contains both B and T cell epitopes with a number of studies demonstrating serial change in HVR1 quasispecies in response to anti HVR1 antibodies (336-338). Conversely, HVR1 demonstrates stability despite serial passage among chimpanzee populations with far fewer non synonymous

substitutions when compared with HVR1 in human hosts(339). This suggests that the changes seen in humans are driven by antibody mediated selection. Studies in humans using immunoglobulin binding techniques have also found differing patterns in HVR1 quasispecies in the antibody bound and unbound fragments(283). The presence of multiple B cell epitopes in HVR1 has formed the theoretical basis for vaccine discovery though the extreme variability of HVR1 has contributed significantly to the disappointing results reported.

1.7 Treatment

1.7.1 Background

The introduction if interferon marked a huge change in the effective treatment of HCV. With efficacy equating to 20-30 % it marked the first significant breakthrough in the successful clearance of HCV. The subsequent addition of ribavirin, a broad acting anti viral resulted in further improvements in treatment success, as did the alteration of interferon into a long acting pegylated form which reduced the requirement of injections from three times per week to weekly. Treatment response was genotype dependent with successful treatment rates of 70-80% for genotypes 2 and 3 and 30-40% for genotypes 1 and 4. Combined pegylated interferon with ribavirin had remained the mainstay of HCV treatment until 2011 when the first protease inhibitors specifically designed to target HCV proteases became available. Boceprevir and telaprevir both demonstrated a significant improvement in treatment efficacy when combined with the previous regime in genotype 1 patients, increasing successful treatment rates to in excess of 70%. However, a number of adverse outcomes in patient with advanced liver disease combined with a significant adverse effect profile have limited their use and the licensing of a number of new direct acting anti virals with greater efficacy has resulted in the withdrawal of both first generation protease inhibitors from market(340). Between 2013 and 2015 a number of new direct acting anti virals have come to market with remarkable improvements in SVR rates for all genotypes regardless of the degree of underlying liver disease, co infection with HIV or transplant status. These drugs have revolutionised the management of hepatitis C and overcome all previously identified viral and host factors associated with lower SVR rates. To date sofosbovir (an NS5B polymerase inhibitor), ledipasvir (NS5A inhibitor), ombitasvir (NS5A inhibitor), paritaprevir (NS3-4A inhibitor), ritonavir (an anti retroviral medication that is included with ombitasvir, paritaprevir and dasabuvir), dasabuvir (NS5B inhibitor), and daclatasvir (NS5A inhibitor) have been licensed for use with combination therapy with or without ribavirin recommended for all genotypes(24).

1.7.2 Treatment Individualisation

During the interval between the emergence of dual therapy and the licensing of the first generation protease inhibitors a number of viral and host factors that are associated with likelihood of SVR were described:

1.7.3 Viral Predictors of SVR using dual therapy with Pegylated Interferon and Ribavirin

1.7.3.1 Hypervariable Region 1 Complexity

Author	Year	Country Study Undertaken	No of Patients	Genotype	Method used to Estimate Complexity	Treatment Received	Duration of Treatment (weeks)	Low HVR1 Complexity Associated with SVR
Okada et al	1992	Japan	6	1	SSCP	IFN β	6-16	Yes
Nakazawa et al	1994	Japan	14	1	SSCP	IFN α	16	No
Koizumi et al	1995	Japan	42	2 (38) 1(4)	SSCP	IFN α	12-24	Yes
Moribe et al	1995	Japan	25	1	SSCP	IFN α	12	Yes
Pawlotsky et al	1998	France	114	1 (65) 3 (21) Others (15)	SSCP	IFN α	12-24	Yes
Lopez-Labrador et al	1999	Spain	122	1	SSCP	IFN α	48	No
Grahovac et al	2000	Croatia	12	1	SSCP	IFN α/Riba	48	Yes
Hino et al	2000	Japan	53	1(27) 2(26)	SSCP	IFN α	24 48	Yes
Yeh et al	2002	Korea	30	1 (24) 2 (6)	Cloning/Shannon Entropy	IFN α	24	Yes
Sandres et al	2000	France	26	1 (2) 2 (8) 3 (16)	Cloning/Shannon Entropy	IFN α	48 24 24	No
Abbate et al	2004	Italy	28	1b	Cloning/Shannon Entropy	PEG IFN α/Riba	48	No
Chambers et al	2005	United States	29	1	Cloning/Shannon Entropy	PEG IFN α/Riba	48	Yes
Moreau et al	2008	Ireland	10	3a	Cloning/Shannon Entropy	PEG IFN α/Riba	24	Yes

Table 1

Summary of previous studies investigating associations between HVR1 complexity and response to HCV treatment.

Quasispecies complexity as has been previously discussed is a measure of the number of unique sequences and their prevalence within a group of clones. HVR1 QS complexity has long been associated with likelihood of treatment success though some findings have been contradictory:

Early studies investigating complexity involved single strand conformation polymorphism technique (SSCP), which was a 3 dimensional agar electrophoresis that was able to separate PCR products with as little as a single nucleotide difference between them. Results using SSCP have been conflicting with Okada et al. first identifying an association between low complexity and SVR among six patients treated with interferon α(341), though no such association was found by Nakazawa et al. in a study of 14 patients(342). In a number of larger scale studies, Koizumi et al. and Moribe et al. demonstrated an association between lower complexity (though in the study the authors use the term diversity) and

treatment response among 42 mostly genotype 2 and 25 genotype 1 patients respectively, a finding that was further corroborated by Grahovac et al. among 12 patients of unknown genotype(343-345). Pawlotsky et al. identified, again using SSCP among 114 patients infected with a variety of genotypes, that low HVR1 complexity was associated with increased SVR rates when treated with interferon alone(346). Yeh et al. reported a similar finding among 26 patients with genotype 1 infection treated with interferon α(347). Lopez-Labrador et al. in examining the complexity among 122 patients with genotype 1 infection did not find an association between baseline HVR1 complexity and SVR(348). Notably, however, these patients were treated for a maximum of six months with interferon α alone, which may have distorted the true effect of complexity on treatment outcomes.

With the more widespread use of genome sequencing, calculation of QS complexity evolved and Sandres et al. studied 13 responders and 13 non responders containing multiple genotypes using Shannon entropy to calculate complexity from 20 cloned sequences per sample and failed to find an association between complexity and SVR(349). Subsequently, Moreau et al and Chambers et al using Sanger sequencing of cloned plasmid samples demonstrated a similar association among 10 patients with genotype 3a infection and 29 genotype 1 patients who underwent treatment with dual pegylated interferon and ribavirin treatment(350, 351). However, Abbate et al. reported no association between HVR1 complexity and SVR using cloned samples but it has been pointed out that they reported complexity at a nucleotide level which contrasts with the studies by Chambers et al. and Moreau et al. where complexity at amino acid level was used(352).

1.7.3.2 HVR1 Diversity

Although, a number of studies in the late nineties suggested an association between low HVR1 QS diversity and SVR, these studies exclusively used SSCP analysis which is more akin to HVR1 QS complexity. The emergence and greater accessibility of cost effective genetic sequencing methods resulted in the emergence of QS diversity as an additional metric which came into widespread use in the description of QS. Diversity came to be defined as the average number of nucleotide substitutions between sequences in a group of clones. Diversity itself could then be adjusted in accordance with the underlying evolutionary model that best represented the pattern of sequence change described by a group of clones(353). A reduction in HVR1 QS diversity is associated with spontaneous clearance of acute Hepatitis C virus infection(308). Furthermore, an early reduction in HVR1 QS diversity in patients on treatment with dual therapy is associated with increased SVR rates(308). Pre treatment HVR1 diversity as a tool to predict likelihood of treatment response has been investigated in a number of studies with variable results. Chambers et al. identified an association between low HVR1 diversity and response to treatment but not SVR among 29 genotype 1 patients who underwent 48 weeks of treatment with pegylated interferon α and ribavirin(351, 354). Fan et al. demonstrated the opposite

in a study of 153 genotype 1 patients, where an Early Virologic Response (EVR) which has been shown to be a marker of SVR was associated with high HVR1 diversity. Notably, this study did not provide data with regard to SVR and the degree of diversity required at 0.53 was by the authors own admission extremely rare. Indeed, although >80% with a HVR1 diversity greater than 0.53 achieved EVR, the limited proportion of patients with such high HVR1 diversity meant that most of the patients who achieved SVR actually had HVR1 diversity <0.53(354).

1.7.3.3 Interferon Sensitivity Determining Region (ISDR)

An association between amino acid substitutions in the non structural protein 5A (NS5A) and SVR was first identified among a cohort of 84 Japanese patients with genotype 1b infection who were treated with interferon α for six months. Patients with wild type amino acid sequences for positions 2209-2248 when compared with HCV-J a consensus sequences for genotype 1b experienced universal treatment failure on interferon therapy. SVR rates for patients with 1-3 and 4-11 amino acid substitutions were 13% and 100% respectively. This section of NS5A was subsequently named the interferon sensitivity determining region (ISDR)(211, 355). This finding was subsequently confirmed in further Japanese, Thai and Chinese cohorts of genotypes 1, 2, and 6 patients(356-358). Notably, while many studies have shown an association between ISDR substitutions and SVR in the Far East, despite numerous studies in Europe and the United States, no similar association has been described though a meta analysis of published data has suggested that this may be due to differences in the underlying prevalence of such mutant ISDR sequences in different geographical regions (359-362). No similar association between ISDR substitutions and response to interferon therapy has been identified to date (363, 364).

1.7.3.4 HCV Core substitutions

Amino acid substitutions at Core70 and Core 91 are associated with reduced response to peygylated interferon and ribavirin among genotype 1b patients(365). Replacement of arginine with glutamine or histidine at position 70 and or the substitution of methionine for leucine at position Core 91 resulted in a higher rate of non response to treatment among 50 Japanese patients. Further studies in 313 and 361 genotype 1b patients in Japan have demonstrated at least a doubling of the rate of developing hepatocellular carcinoma in patients who are infected with viruses containing these mutant Core substitutions(366, 367). These substitutions appear to confer resistance to treatment that is somewhat, but not completely overcome by prolonging treatment to 72 as opposed to 48 weeks(368). Such mutant Core amino acid substitutions appear to increase interferon resistance by up regulating interleukin-6 (IL-6), which in turn increases suppressor of cytokine signalling 3 (SOCS3)(369). A similar association between core amino acid substitutions has been identified in genotype 2 patients, but not to date in genotype 3 patients (370, 371).

1.7.4 Host Predictors of SVR using Dual Therapy with Pegylated Interferon and Ribavirin

1.7.4.1 Interferon γ inducible protein 10 (IP 10)

IP 10 first became implicated in characterising the host immune response to HCV when it was noted that IP 10 levels correlated with histologic disease severity in chronic infection(372). Associated with T lymphocyte activation, IP 10 is a pro inflammatory chemokine. An association between IP 10 levels and SVR was almost simultaneously described by two groups in Spain and Sweden. Diago et al. identified that IP 10 levels were much lower among responders among 103 patients with chronic genotype 1 HCV infection who were treated with pegylated interferon and ribavirin and suggested that this could be used to predict SVR(373). Lagging et al. demonstrated that low IP 10 levels, even among patients with high body mass index or low viral loads (both negative predictors of SVR), were associated with higher rates of SVR. An IP 10 cut off of less than 150 pg/mL for optimal response and between 150-600pg/mL for improved response was described while those patients who had an IP 10 level >600pg/mL demonstrated a very poor response to pegylated interferon and ribavirin(374, 375). Low IP 10 levels at treatment induction are also associated with rapid reductions in HCV viral load among genotype 1,2 and 3 patients, a feature which has also been shown in patients co infected with HIV(376, 377). Higher IP 10 levels have also been associated with more advanced liver fibrosis(378). As a pro inflammatory chemokine, which acts by attracting activated lymphocytes, the association between high levels and poor response had remained difficult to reconcile. It has recently been proposed that the form of IP 10 produced in these patients may in fact be a modified antagonistic form of IP 10, which results in dysregulation of the host immune response(379).

1.7.4.2 IL 28 polymorphisms

1.7.4.2.1 Rs12979860 Genotype 1

Polymorphisms of the IL28 gene which is involved in the regulation of endogenous interferon signalling are also associated with response to dual therapy. Ge et al. first noted an association between allele pairs at rs12979860 and response to pegylated interferon and ribavirin among a cohort of 1,600 genotype 1 infected treatment naive patients. Patients with a CC at this site had a twofold higher response rate to treatment when compared with those who had CT or TT, though a subsequent meta analysis of 10 papers describing this phenomenon has suggested that the odds ratio of treatment response is 5.52 (3.74-8.15) (380, 381). This allele codes, located 3kb upstream from the IL28β gene which codes for interferon λ, part of the family of interferons which are known to have broad antiviral activity(382). An association between responder substitutions at rs 12979860 and spontaneous clearance of HCV was also described among 1008 genotype 1 infected patients, 388 of whom had spontaneously cleared their infection(383). Interestingly carriers of responder rs12979860 C/C who

go on to develop chronic infection have been shown to have higher viral loads and more histologic evidence of severe lobular inflammation and fibrosis progression across all genotypes(384, 385).

1.7.4.2.2 Rs12979860 Genotype 3

While there have been many studies investigating the clinical implication of rs12979860 in genotype 1, studies specifically evaluating patients with genotype 3 infection are fewer and have described more heterogeneous results. A meta-analysis of data from five studies gave an odds ratio of 1.23 (.071-2.14) for SVR suggesting no association though it should be noted that a three of these studies combined genotype 2 and 3 patients(381). To date four studies have elucidating treatment response to IL28 polymorphisms have been published. Two European studies identified an association between C/C and rapid decrease in viral load but found no association with SVR(382, 386). Two other studies in the Indian subcontinent that included 356 and 105 patients have identified an association between C/C and SVR among genotype 3 patients. We note however that the SVR rates for C/T at 30.5% and 56.4% and T/T at 8.4% and 22.2% were extremely low(387, 388). Finally, one European study including 267 patients which was not included in the earlier meta-analysis has suggested an association between C/C and SVR, but this study combined genotype 2 and 3 patients(389).

1.7.4.2.3 rs8099917 Genotype 1

Further studies including a seminal paper by Rauch et al. investigating IL28 using multivariate logistic regression identified a second site rs8099917 which locates 8kb downstream of IL28 and codes for interferon λ2(382, 390). This site was also associated with higher rates of spontaneous clearance and interferon based treatment success particularly in genotype 1 and 4 patients. TT (spontaneous clearance 76%, SVR 68%) at this site was characterised as optimal with lesser rates of spontaneous clearance and interferon response among patients with GT (spontaneous clearance 22%, SVR 29%) and GG (spontaneous clearance 1%, SVR 3%)(390, 391). The identification of these polymorphisms in the interferon λ coding genes and its association with spontaneous HCV clearance suggests an important role for innate immunity in viral clearance. One large meta-anaysis of data has shown marked improvement in SVR among patients with genotype 1 infection (odds ratio 4.28 (2.87-6.38))(381).

1.7.4.2.4 rs8099917 Genotype 3

Meta-analysis of five early studies investigating SVR among patients with differing rs8099917 polymorphisms did not show a statistically significant association with SVR (odds ratio 1.4 (0.98-2.00), but this included studies where genotype 2 and 3 were combined(381). Three subsequent studies on genotype 3 patients in Europe (one of which combined genotype 2 and 3) have shown an association between polymorphisms at this site and rapid response to treatment but not with SVR(382, 386, 389).

Aziz et al. reported similar findings among patients with T/T at this site among 105 patients in Pakistan(387). Patients with TT genotypes appear to have a more progressive disease with more advanced fibrosis and increased risk of hepatocellular carcinoma(392).

1.7.4.2.5 Other Single Nucleotide Polymorphisms (SNPs)

The use of genome wide association studies has identified a number of other SNPs in IL28 that have been associated with SVR following dual therapy with pegylated interferon and ribavirin, though they have not to date been studied in the same detail as the two previously discussed. The SNP that has been studied in greatest detail is rs12980275. This SNP is associated increased rates of RVR in genotypes 1, 2, and 3 and SVR in genotype 1 when trial data from 253 Caucasian individuals who were enrolled in the HCV-DITTO trial was retrospectively reviewed(393, 394). With the development of massive parallel sequencing techniques it has been possible to identify up to 18 IL28 SNPs that are associated with SVR in genotype 1 patients.

However, the development of highly efficacious direct acting antiviral drugs has overcome the association between IL 28 polymorphisms and poor treatment response (21, 395).

1.8 Conclusion

The individualisation of HCV treatment has been largely on the basis of population parameters. HCV genotype has been widely shown to affect treatment prognosis as has race, body mass index, alcohol consumption, advancing liver disease and cirrhosis.

A number of retrospective studies have identified associations between HVR1 quasispecies parameters and likelihood of successful treatment response with dual therapy. The natural history of short interval change in HVR1 quasispecies is absent from the literature. The degree to which these changes are the result of antibody escape is unclear.

Recent studies have also identified a number of host and viral molecular markers that might assist in identifying patients who are likely to achieve sustained viral response. Pre treatment identification of success is important for a number of reasons:

1. Treatment with interferon and ribavirin is associated with significant side effect related morbidity including fatigue, anaemia, thyroid dysfunction, depression and increased suicide risk.
2. HCV treatment is resource and cost intensive and optimising success rates has potentially significant cost saving implications for health care systems.
3. The new highly efficacious DAAs are extremely costly and the potential to identify candidates with high response rates to dual therapy would allow healthcare professionals to target these costly drugs at the patients with poorer outcomes, thus allowing the treatment of more patients with finite resources.

We studied short interval changes in HCV HVR1 quasispecies parameters and both host and viral predictors of response to treatment in order to investigate for treatment individualisation strategies to optimise sustained virologic response rates. We also separated the immunoglobulin bound from immunoglobulin free HVR1 sequences to investigate whether quasispecies change was antibody driven.

As the study progressed, the composition of our cohort precluded the exploration and discovery of treatment individualisation strategies. We therefore focused on describing in depth the evolution of HCV over short time intervals in a novel cohort of chronically infected individuals. This facilitated the identification of differing patterns of HCV change depending on the degree of underlying liver fibrosis. By using molecular clock techniques we then explored for evidence supportive of a variable fidelity polymerase which we hypothesize emerges in a phase specific fashion as the virus adapts to the host immune system. We investigated our samples using Sanger sequencing of nested polymerase chain

reaction (PCR) product of the E1E2 section of the virus envelope (a segment that includes HVR1. These were compared with Sanger sequences generated from amplified plasmid clones. Our PCR product (320 base pairs) was of a suitable lenght for Sanger sequencing. We identify potential pitfalls in the description of an underlying QS using amplicon Sanger sequencing.

Finally, we used 454 next generation sequencing to interrogate the QS in greater depth. This allowed comparison between all three descriptive strategies. 454 sequencing with the use of appropriate screening tools can provide information with regard to minor components within the QS which is not achievable using cloning strategies. However, 454 next generation sequencing has its own limitations notwithstanding its high cost. The technology we used requires post hoc reconstruction of the underlying sequences. This can potentially distort the QS and as a result sequences generated that form <0.1% of the underlying sample are removed. Finally, 454 sequencing produces extremely large volumes of data and requires rigorous bioinformatic approaches in order to provides accurate interpretation which is time and labour intensive.

Chapter 2

Methods

2.1 Methods

2.1.1 Aims

As outlined in the introduction the purpose of the project was to describe short interval change in HVR1 quasispecies and investigate whether QS complexity and diversity (parameters that have previously been shown to correlate with treatment outcomes) could be utilised in future models to optimise response to pegylated interferon and ribavirin.

Using nested PCR and cloning strategies, followed by DNA amplification and sequencing of a plasmid encoded 320 base pair section of the envelope protein encompassing HVR1 we planned to generate 10-24 clones for each sample collected. Individual cloned sequences were provided in the form of raw fasta (.fas) files and electropherograms by the sequencing company (Eurofins DNA). All fasta sequences were compared with the corresponding electropherogram manually to correct any mis reads. All individual sequences for a single sample were combined in a single fasta file and visualised using MEGA6. Sequences that included short insertions (<10bp) between the plasmid primers were removed. The remaining sequences were then aligned using ClustalW. All resuyultant sequences were 320 bp in length. The base pairs corresponding with the forward and reverse primers were removed from downstream analysis as their inclusion could potentially distort the accuracy of the data. Once the sequences had been aligned and inspected we planned to calculate HVR1 complexity and diversity for each sample and the degree of HVR1 change or divergence both between groups of clones and from the original group of clones to investigate for patterns of HVR1 change comparing the data with the published literature.

Cloning data outcomes was visualised using unrooted phylogenetic trees, a method that uses specific evolutionary models to estimate sequences origins using bifurcating trees, such that the tree can provide an estimate as to how the group of clones is likely to have developed from an unknown common ancestor. For the analysis of our cloning data we identified the optimal evolutionary model using jmodeltest as Generalised Time Reversible with Invariant sites and a gamma distribution (GTR+I+G).

We performed an initial tree exploration for basic errors and to out rule contamination of samples we using a Neighbour Joining model with a boot strap of 500 replicated for the purposes of tree optimisation as this provided a quick overview of the data. Early exploration of the outputs from Neighbour Joining methods identified problems with the density of HVR1 substitutions and difficulties that the program had with identifying identical sequences and placing them together on the tree. As a result we included only unique nucleotide sequences from each sample for future tree exploration.

Gamma distribution was calculated individually for each individuals group of clones. We then compared trees generated using GTR+I+G with maximum composite likelihood trees including the patient specific gamma distribution and found no significant difference between these models but the computational requirement of the former led us to choose the latter for our cloning study. Difficulties with tree optimisation were overcome by increasing bootstrap replicates to 10,000 per tree.

During the period of the study a seminal paper on long term HVR1 change was published describing groups of clones as subpopulations and establishing that these are potentially maintained over years of chronic infection, while simultaneously describing HVR1 change between clades on a phylogenetic tree as time ordered phylogenetic change. We therefore also interrogated our data using partitioned analysis for short interval subpopulation dynamics and evidence of time ordered phylogeny.

The time required to generate cloning outputs is significant and was identified as a potential obstacle to the use of these strategies in treatment optimisation. We therefore sought to establish whether nested PCR amplicons can provide adequate accurate information with regard to the underlying HVR1 quasispecies and whether amplicons could be used to predict QS behaviour.

Early analysis of cloning outputs indicated a spectrum of HVR1 change from stasis to widespread time ordered phylogenetic change. In an effort to establish whether these short term changes reflected a sustained or episodic process we included a retrospective group of clones from our sample library for each subject, where a suitable sample was available. We also investigated whether adaptive immune pressures were responsible for the HVR1 changes identified by separating the serum samples into immunoglobulin bound and free fractions and sequencing the resultant amplicons.

Finally, using next generation sequencing, we investigated patterns of HVR1 change at a third level of depth (amplicon, cloning and pyrosequencing). The inclusion of pyrosequencing data necessitated the transfer from maximum likelihood phylogenetic trees to GTR+I+G trees generated as the density of sequence depth required more rigorous evolutionary model application.

2.2 Study Design

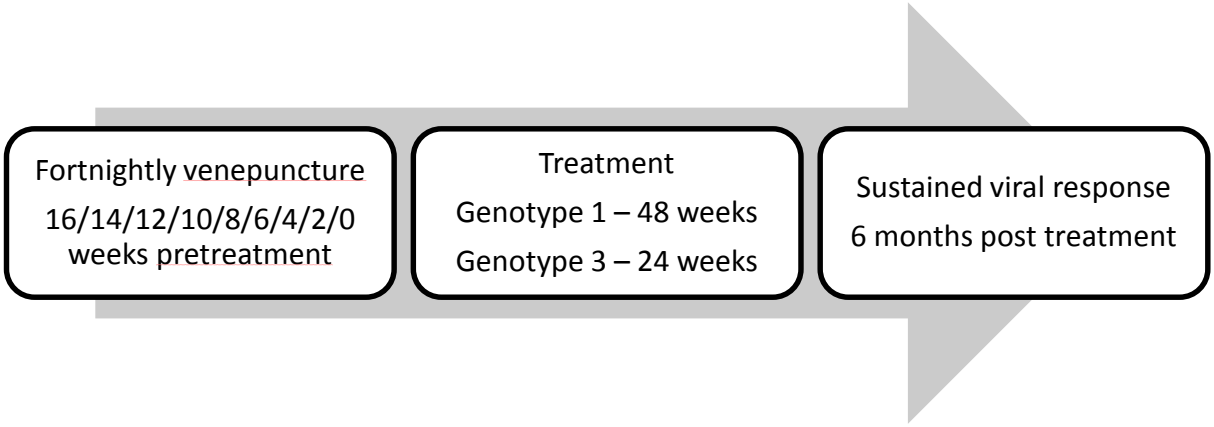


Fig 1 Study Design

2.2.1 Subject Recruitment

Patients who were deemed eligible for treatment at the tertiary referral centre exceeded the capacity of the centre and as a result there was a time lag between subjects being deemed eligible and the date of starting treatment. Accordingly, there was a suitable window for sample collection which did not delay treatment as any delay due to study participation would have been unethical.

Ethical approval for the study was sought and obtained under the auspices of the Clinical Research Ethics Committee of the Cork Teaching Hospitals.

All suitable patients were first approached by a clinician with whom rapport had already previously been established and asked whether they would be willing to discuss entry into the study. I met all willing subjects and outlined the project and the requirement to attend for blood tests that would not have otherwise been required in advance of treatment. All subjects were provided with an outline of the study and all questions were answered with in addition an explanation that study participation was not a prerequisite for treatment and that withdrawal of consent would be possible at any time without repercussion.

In total I approached 28 patients of whom 26 provided informed consent for study participation. Two patients declined study entry, one of whom proceeded to treatment and the second person was not treated as the supervising clinicians had concerns that ongoing intermittent alcohol intake and frequent nonattendance for clinic appointments contraindicated safe treatment.

One patient attended for her first sample 16 weeks prior to commencing treatment and had become pregnant and treatment was postponed until after she finished breastfeeding. A second patient attended erratically for sample collection but did not proceed to treatment as he was deported

following an unsuccessful application for political asylum. Finally, a third patient developed gallstone pancreatitis shortly before he was due to commence sample collection and treatment was postponed until such time as he had recovered from interval cholecystectomy.

2.2.2 Consent Form

SUBJECT INFORMATION SHEET

Protocol Number:_____ Patient Name:_____

Title of Protocol: Do the dynamics of quasispecies complexity and IP-10 concentration in chronic hepatitis C provide an opportunity to individualise treatment strategies?

Principal Investigator:

Dr Orla Crosbie Phone: 021-4922066

You are being asked to participate in a research study. The doctors at University College Cork study the nature of disease and attempt to develop improved methods of diagnosis and treatment. In order to decide whether or not you want to be a part of this research study, you should understand enough about its risks and benefits to make an informed judgment. This process is known as informed consent. This consent form gives detailed information about the research study, which will be discussed with you. Once you understand the study, you will be asked to sign this form if you wish to participate.

Why is this study being run?

It is estimated that Hepatitis C affects up to 170 million people worldwide. You are invited to participate in a research study where we aim to study the hepatitis C virus that could be used to identify improved treatment strategies.

If you agree to participate in the study several blood samples will be obtained over a four month period prior to starting your treatment. The first visit will take place at Cork University Hospital. Participation will involve blood tests every two weeks for four months.

This study will involve up to 40 patients. All of these other subjects will be between 18 and 75 years of age.

Study Procedure

You will be advised of the purpose of the study and the procedures which will be undertaken. You will be given a copy of the subject information sheet, which will explain what is required from you. If interested, you will then be requested to read and sign the Informed Consent form, and receive a signed copy.

A venous blood sample (10ml) will be obtained and the virus building blocks (RNA) will be assessed. The blood sample will also be assessed for an inflammatory protein (IP 10). Eight blood samples will be collected in total every fortnight over a period of four months. We will use these blood samples to monitor for changes in the virus and inflammatory proteins over that period.

What happens if I start the study and change my mind later?

You do not have to take part in the study, participation is entirely voluntary. Refusal to participate, or discontinuing participation at any time, will involve no penalty, loss of benefits or denial of treatment or services by the Cork Teaching Hospital or the participating doctor.

Who is performing the study?

The study is being undertaken by Dr Daniel Schmidt-Martin under the supervision of Dr Orla Crosbie. Dr Schmidt-Martin will be taking the blood samples and is currently studying for a PhD. in the area of Hepatitis C.

Will I experience any unpleasant side effects?

During the collection of your blood sample, you may experience a slight scratch, which may be uncomfortable for a moment but quickly passes.

Funding of trial

There are no cost implications for the Health Board or to you. The management of patients and investigative tests will comply with current standards of care. Cost of research tests will be incurred by University College Cork.

Confidentiality

All the information gathered from this study will be stored on a computer, paper files and will be treated confidentially. You will be identified only by a subject number. In the event of any publication regarding this study, your identity will not be disclosed.

What happens if there is anything I do not understand?

If there is anything you are not sure about, the Doctor will be happy to explain in more detail to yourself or your relatives, guardians (or legal representative of required). The study will be fully explained to you before you decide if you want to take part. If you have any problems or questions after the study has started you may call:

Dr Orla Crosbie

Consultant Gastroenterologist

Cork University Hospital, Cork, Tel 021 4546400

CONSENT BY SUBJECT FOR PARTICIPATION IN RESEARCH PROTOCOL

Protocol Number:

Patient Name:_____

Title of Protocol: Do the dynamics of quasispecies complexity and IP-10 concentration in chronic hepatitis C provide an opportunity to individualise treatment strategies?

Principal Investigator:

Dr Orla Crosbie Phone: 021-4922066

Participation in this study is voluntary and you may withdraw at any time for any reason

The research project and the treatment procedures associated with it have been fully explained to me. All experimental procedures have been identified and no guarantee has been given about the possible results. I have had the opportunity to ask questions concerning any and all aspects of the project and any procedures involved. I am aware that participation is voluntary and that I may withdraw my consent at any time. I am aware that my decision not to participate or to withdraw will not restrict my access to health care services normally available to me. Confidentiality of records concerning my involvement in this project will be maintained in an appropriate manner. When required by law, the records of this research may be reviewed by government agencies and sponsors of the research.

I understand that the sponsors and investigators have such insurance as is required by law in the event of injury resulting from this research.

I, the undersigned, hereby consent to participate as a subject in the above described project conducted at the Cork Teaching Hospitals. I have received a copy of this consent form for my records. I understand that if I have any questions concerning this research, I can contact the doctor(s) listed above. If I have further queries concerning my rights in connection with the research, I can contact the Clinical Research Ethics Committee of the Cork Teaching Hospitals, Lancaster Hall, 6 Little Hanover Street, Cork.

After reading the entire consent form, if you have no further questions about giving consent, please sign where indicated.

Subject's Signature:_____ Date_____

dd mon yy

Name (BLOCK LETTERS):_____ Time:_____

Investigator's signature:_____ Date_____

dd mon yy

Name (BLOCK LETTERS):_____

2.2.3 Patient Attendance

Patient attendance was satisfactory. Early in the project a number of patients did not attend for their second sample and the possibility of arranging for an automated text reminder was discussed but as participation was voluntary and attendance potentially associated with cost for the patient it was agreed that this might constitute undue pressure and may in certain circumstances risk withdrawal from the study. Instead patients were provided with a written confirmation of their next date and time for blood testing after each sample had been taken.

As treatment duration ranged from 24 to 48 weeks and up to 72 weeks among slow genotype 1 responders and was often associated with significant morbidity, patients were encouraged to consider taking a holiday prior in advance. This resulted in a number of missed data points. Furthermore, the study period coincided with the imposition of a no fly zone in Europe following the eruption of a volcano in Iceland and one of the patients became stranded in New York and missed a number of appointments.

	Weeks pre treatment								
Subject	16	14	12	10	8	6	4	2	0
A	Y	Y	N	Y	Y	N	Y	Y	Y
B	Y	Y	Y	N	N	Y	Y	Y	Y
C	Y	Y	Y	N	Y	N	N	N	Y
D	Y	Y	Y	N	Y	N	Y	N	Y
E	Y	Y	Y	N	Y	N	Y	N	Y
F	Y	Y	Y	N	Y	Y	Y	Y	Y
G	Y	Y	N	Y	Y	Y	Y	Y	Y
H	Y	Y	Y	Y	Y	Y	Y	Y	Y
I	Y	Y	Y	Y	Y	Y	Y	Y	Y
J	Y	Y	Y	Y	Y	Y	Y	Y	Y
K	N	Y	Y	Y	Y	Y	Y	Y	Y
L	Y	Y	N	Y	Y	Y	Y	Y	Y
M	Y	N	Y	Y	Y	Y	Y	N	Y
N	Y	N	N	N	N	N	N	Y	Y
O	Y	Y	Y	N	Y	Y	Y	N	Y
P	Y	N	Y	Y	Y	Y	Y	Y	Y
Q	Y	Y	Y	Y	Y	N	Y	Y	Y
R	Y	Y	Y	Y	Y	Y	Y	N	Y
S	Y	Y	Y	N	Y	Y	Y	N	Y
T	Y	Y	Y	N	Y	N	Y	Y	Y
U	N	Y	Y	Y	Y	Y	Y	Y	Y
V	N	Y	Y	Y	Y	Y	Y	Y	Y
W	Y	Y	Y	Y	Y	Y	Y	N	Y

Fig 2. Patient attendance

2.3 Methods in Chapter 4

Subjects attended for venepuncture every two weeks for a period of 16 weeks prior to commencing treatment with pegylated interferon and ribavirin. All subjects were prospectively recruited in an unselected fashion with inclusion criteria confined to chronic hepatitis C infection between the ages of 16 and 75. Entry into the study was voluntary and no compensation financial, or otherwise, was provided to study participants. All participants were attending outpatients in a tertiary referral centre in the Republic of Ireland for ongoing management of chronic hepatitis C.

Subjects provided written informed consent and the study was undertaken under the governance of the Clinical Research Ethics Committee of the Cork Teaching Hospitals. Samples were centrifuged within 2 hours, and stored at -70 C within six hours of collection.

2.3.2 RNA extraction and amplification

The extraction of HCV RNA was performed by use of the Total Nucleic Acid Isolation protocol on the MagNA Pure LC (Roche Diagnostics Ltd., UK) automated platform.

2.3.3 Reverse Transcription

Reverse transcription was commenced using 8 µl of the isolated RNA and 0.5 µg random primers (Promega, Madison, WI). This mixture was heated at 75°C for ten minutes followed by a brief incubation on ice. cDNA synthesis was performed at 42 ° C with a mastermix containing 400 µM dNTPs (Roche, UK), 40 units RNAse inhibitor (Promega, Madison, WI), 4 µl of AMV RT 5x buffer and 10 units of AMV reverse transcriptase (Promega, Madison, WI) in a final volume of 20 µl for 60 minutes followed by 3 minutes at 94 °C for enzyme denaturation.

2.3.4 Nested PCR of E1/E2

The E1/E2 segment including the HVR1 were amplified using nested PCR producing a 320 bp fragment corresponding to nucleotides 1254 to 1572 of reference strain HCVGENS1 genotype 3a (Genbank : X76918). Pwo polymerase was used for the amplification of blunt ended products for cloning.

Nested PCR protocol:

Primary PCR

Forward 5'- ATG GCA TGG GAT ATG AT -3'

Reverse 5'- AAG GCC GTC CTG TTG A -3'

A 50 µl master mix was produced combining 5.0µl RT PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 3.0µl $MgSO_4$, 32.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 15 s, 51°C for 30 s, 72°C for 30 s and a final extension at 72°C for 7 minutes.

Secondary PCR

Forward 5'- GCA TGG GAT ATG ATG ATG AA -3'

Reverse 5'- GTC CTG TTG ATG TGC CA -3'

A 50 µl master mix was produced combining 4.0µl primary PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 34.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 15 s, 53°C for 30 s, 72°C for 30 s and a final extension at 72°C for 7 minutes.

2.3.5 PCR amplicon of the correct size was confirmed using gel electrophoresis.

2.3.5.1 2% agarose gel preparation

1. Measure out 2g of agarose.
2. Pour agarose powder into microwavable flask along with 100mL of 1xTAE
3. Microwave for 1-3min (until the agarose is completely dissolved and there is a nice rolling boil).
4. Let agarose solution cool down for 5min.
5. Pour the agarose into a gel tray with the well comb in place.

Note: Pour slowly to avoid bubbles which will disrupt the gel. Any bubbles can be pushed away from the well comb or towards the sides/edges of the gel with a pipette tip.

Place newly poured gel at 4°C for 10-15 minutes OR let sit at room temperature for 20-30 minutes, until it has completely solidified.

2.3.5.2 Loading Samples and Running an Agarose Gel:

1. Add loading buffer to each of your digest samples.

   Note: Loading buffer serves two purposes: 1) it provides a visible dye that helps with gel loading and will also allows you to gauge how far the gel has run while you are running your gel; and 2) it contains a high % glycerol, so after adding it your sample is heavier than water and will settle to the bottom of the gel well, instead of diffusing in the buffer.
2. Once solidified, place the agarose gel into the gel box (electrophoresis unit).
3. Fill gel box with 1xTAE (or TBE) until the gel is covered.
4. Carefully load a molecular weight ladder (100-1,000bp) into the first lane of the gel.
5. Carefully load your samples into the additional wells of the gel.
6. Run the gel at 80 until the dye line is approximately 75-80% of the way down the gel.
7. Turn OFF power, disconnect the electrodes from the power source, and then carefully remove the gel from the gel box.
8. Using any device that has UV light, visualize your DNA fragments.

2.3.5.3 Recipe for TAE

1. One liter 50X stock of TAE Tris-base:
2. 242 g Acetate (100% acetic acid): 57.1 ml
3. EDTA 100 ml 0.5M sodium EDTA
4. Add dH2O up to one litre.

To make 1x TAE from 50X TAE stock, dilute 20ml of stock into 980 ml of DI water

When running gels negative controls were included in parallel in order to observe for possible cross contamination of samples.

2.3.6 Cloning

Cloning was performed using Zero Blunt® TOPO® PCR Cloning Kit using chemically competent cells (TOP10) (Invitrogen, Belgium).

Cloning reaction:

PCR product 4µl

Salt solution 1µl

TOPO® vector 1µl

1. Mix reagents gently and incubate for 5 minutes at room temperature (22-23°C).
2. Place the reaction on ice and proceed to transformation of competent cells.

Transformation of Competent Cells

3. Add 2 µL of the TOPO® Cloning reaction from Perform the TOPO® Cloning reaction into a vial of One Shot® chemically competent E. coli and mix gently. Do not mix by pipetting up and down.
4. Incubate on ice for 5–30 minutes.
5. Heat-shock the cells for 30 seconds at 42°C without shaking.
6. Immediately transfer the tubes to ice.
7. Add 250 µL of room temperature S.O.C. medium (provided in cloning kit).
8. Cap the tube tightly and shake the tube horizontally (200 rpm) at 37°C for 1 hour.
9. Spread 10–50 µL from each transformation on a prewarmed selective plate and incubate overnight at 37°C. To ensure even spreading of small volumes, add 20 µL of S.O.C. medium We recommend that you plate two different volumes to ensure that at least one plate will have well-spaced colonies.
10. Two plates were prepared for each cloning reaction to increase the likelihood of well space colonies.

2.3.7 Preparing Agar plates

Making the LB Agar

1. Add 250 mL of dH2O to a graduated cylinder.
2. Weigh out 20g of premix LB Agar powder (VWR)

3. Add dH2O to total volume of 500 mL and transfer to 1 L flask
4. Put on stirring hot plate and heat to boil for 1 min while stirring.
5. Transfer to 1 L pyrex jar and label with autoclave tape.
6. Autoclave at liquid setting for 20 minutes in a basin making sure to loosen top
7. Let agar cool to ~55C (you should be able to pick up the jar without a glove)

Pouring Agar Plates

1. Make sure bench top has wiped down with bleach/EtOH.
2. Pour a thin layer (5mm) of LB Agar (~10mL) into each plate being careful to not lift the cover off excessively (you should be able to just open up enough to pour). Swirl plate in a circular motion to distribute agar on bottom completely.
3. Let each plate cool until its solid (~20 minutes) then flip so as to avoid condensation on the agar.
4. Store plates in plastic bags in fridge with: name, date and contents (note any additive).

2.3.8 Amplification of Cloning Plasmid

1. Performed using illustra™ Templiphi 100 Amplification Kit (GE Healthcare Freiburg, Germany) as follows:
2. Add 5µl aliquots of sample buffer to each well in a microwell plate.
3. A small portion of an individual colony was added to the sample buffer taking care not to transfer agar.
4. Seal the microwell plate and denature at 95°C for 3 minutes and cool the samples to 4°C.
5. Prepare a master mix for the number of colonies to amplified containing 5µl of reaction buffer and 0.2µl of enzyme mix.
6. Add 5µl of master mix to each sample well in the cooled denatured product from step 3.
7. Incubate the supernatant at 30°C for 10-14 hours (the kit guidelines suggest 4-18 hours but we did not perform the reaction at either extreme) and inactivate the enzyme to complete the process by incubating at 65°C for 10 minutes.

2.3.9 Preparation of samples for sequencing

1. Add 3µl of templiphi amplification product to 12µl of sample buffer.
2. Seal microwell and label with sample identifier and seal the microwell within a zip lock bag.
3. Send to sequencing in a padded envelope to Eurofins MWG Operon (Ebersberg, Germany) using M13 reverse priming site as the forward primer.

2.3.10 Sequence analysis

1. Sequence similarity was compared using the BLAST web based tool http://blast.ncbi.nlm.nih.gov/Blast.cgi.
2. The sequences were aligned using CLUSTALW and analysis was performed following the exclusion of sequences which were either incomplete or contained stop codons.
3. The optimum evolutionary model for analysis was determined using jModeltest.
4. Intra sample genetic diversity was calculated using the generalised time reversible model (GTR+I+G).
5. Sequence divergence was calculated using a gamma distributed maximum likelihood evaluation pairwise genetic distance between the groups of clones from each sample.
6. Genetic complexity was described using normalised Shannon entropy [NSE] which was calculated as follows: $Sn = \sum_i[(p_i \times \ln p_i)\ln n]$, where p_i is the number of times each particular sequence appears in the QS and n is the number of sequences in the sample.
7. Phylogenetic analysis was performed using MEGA5 maximum composite likelihood with a bootstrap value of 10,000.
8. Codon specific selection pressures were estimated using Random Effects Likelihood (REL) and evidence for sequence wide selection was established using a PARtitioning approach for Robust Inference of Selection (PARRIS) through the www.datamonkey.org server.
9. Intra host virus population evolution was further visualised in the form of median joining networks (MJN) using SplitsTree4.
10. PAQ was used to identify subpopulations with a minimum difference of 15% at amino acid level between all subpopulations. PAQ is a software suite which can partition sequences into groups using genetic distances either at nucleotide or amino acid level (www.vetmed.iastate.edu/units/carplab/). Initial screening of the data required exploration of the optimum radius using the Net weight evaluation method. Each cluster was subsequently evaluated for sub groups by using the sub group analysis menu. Once the optimum radius had been found the average Hamming Distance from the central sequence within the cluster was calculated as a measure of how compact the swarm was around the central sequence. We found an amino acid radius of 4 changes identified distinct subpopulations within the dataset in all subjects.

2.3.11 Calculation of nucleotide substitution rates

1. Nucleotide substitution rates per nucleotide site per year and the ratio of substitutions at codon position 1+2 to codon position 3 were calculated using MCMC analysis on Beast.
2. Fasta files were converted to Nexus files

3. Nexus files were imported into BEAUTi v1.8.1
4. The files were not partitioned
5. The date of the sequences relative to each other were assigned individually
6. Following a number of trail runs I found HKY provided the best fit model and used estimated base frequency.
7. In order to compare substitution rates in codon positions 1+2 with position 3 I partitioned the sequences and calculated substitution sites for both separately. An overall substitution site was also calculated on a repeat run.
8. I used a strict clock to estimate the substitution rates as this provided the best fit outcomes.
9. The initial clock rate was specified as 1×10^{-3} as this provided the best fit for early optimisation of estimated substitution rates.
10. The length of chain was specified as 10,000,000 with an initial burn in of 1,000,000 and optimal trees were logged every 1,000 operations.
11. Analysis was performed by importing the BEAUTI file into BEAST v1.8.1.
12. The output files were interpreted using Tracer_v1.5 by visualisation of the normality of the operation output and deemed satisfactory if estimated sample size was greater than 200.

2.3.12 Statistical Analysis

1. Continuous variables were tested for normality using a Kolmogorov-Smirnov test.
2. Analysis of QS continuous variable metrics was performed using Mann Whitney U as the data was not normally distributed
3. Categorical variables were analysed using Chi squared analysis.

2.4 Methods in Chapter 5

Serum was collection and storage as previously described.

RNA extraction, reverse transcription, *Pωo* nested PCR of 320 bp E1/E2 product, cloning, amplification, and sequencing as previously described.

2.4.1 Taq Polymerase nested PCR

Nested PCR protocol:

Primary PCR

Forward 5'- ATG GCA TGG GAT ATG AT -3'

Reverse 5'- AAG GCC GTC CTG TTG A -3'

A 50 µl master mix was produced combining 5.0µl RT PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 10.0µl 10x PCR buffer, 3.0µl $MgCl_2$, 27.5µl H_2O, and 0.5µl Go*Taq* DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 15 s, 51°C for 30 s, 72°C for 30 s and a final extension at 72°C for 7 minutes.

Secondary PCR

Forward 5'- GCA TGG GAT ATG ATG ATG AA -3'

Reverse 5'- GTC CTG TTG ATG TGC CA -3'

A 50 µl master mix was produced combining 4.0µl primary PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 10.0µl 10x PCR buffer, 2.0µl $MgCl_2$, 29.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 15 s, 53°C for 30 s, 72°C for 30 s and a final extension at 72°C for 7 minutes.

PCR Product was run on a 2% agarose gel with negative controls in parallel.

2.4.2 PCR product Purification

1. QIAquick PCR Purification Kit© (Qiagen, Venlo, Netherlands)
2. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.
3. If pH indicator I has been added to Buffer PB, check that the color of the mixture is yellow. If the color of the mixture is orange or violet, add 10 µl of 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.
4. Place a labelled QIAquick spin column in a provided 2 ml collection tube.

5. To bind DNA, apply the sample to the QIAquick column and centrifuge at 17,9000 g for 30–60 s.
6. Discard flow-through. Place the QIAquick column back into the same tube.
7. To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30–60 s.
8. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min. IMPORTANT: Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.
9. Place QIAquick column in a clean labelled 1.5 ml microcentrifuge tube.
10. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water (pH 7.0–8.5) to the center of the QIAquick membrane and centrifuge the column for 1 min.
11. IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA. The average eluate volume is 48 µl from 50 µl elution buffer volume, and 28 µl from 30 µl elution buffer.
12. The resultant product can be run on a gel, sent for sequencing, or stored at either 4°C or frozen.

2.4.3 Sequence analysis

The resultant sequences were examined for evidence of stop codons and the electropherograms were visually inspected for inaccurate automated nucleotide base reads.

The amplicon sequences for HVR1 were then combined with all unique nucleotide HVR1 sequences generated for each sample for the same study subject. The combined multiple sequence alignment was transferred into Mega 5 and a phylogenetic tree calculated using a maximum composite likelihood neighbour joining tree with a boot strap replicate value of 10,000 for the purposes of tree optimisation. The amplicon samples for each individual were then labelled on the overall tree.

Additionally, we generated multiple sequence alignments for the amplicon sequence combined with the unique nucleotide HVR1 clones generated from the equivalent sample and visualised the amino acid sequences using multalin (http://multalin.toulouse.inra.fr/multalin/).

Statistical analysis

Continuous variables were compared using Pearson correlation and Mann Whitney U difference of two medians.

2.5 Methods in Chapter 6

We performed the identical reverse transcription, nested pcr, cloning, sequence amplification and sequencing as had been performed on our prospective samples.

The sequences generated were examined, screened for errors and analysed as per the previous samples.

We examined the library for the presence of the subpopulations that had been previously described and investigated for novel subpopulations.

HVR1 QS divergence was calculated and the overall divergence from the original sample was compared to each sample within the prospective study in order to investigate for either divergence or convergence of HVR sequences.

Finally, we used Beast to estimate HVR1 substitution rates and compared these to the rates described over short intervals in our prospective samples.

2.6 Methods in Chapter 7

2.6.1 Ultradeep pyrosequencing data generation, handling and error correction.

Amplicons were quantified using a Biophotometer (Eppendorf) and diluted to a final concentration of 1×10^7 molecules/ml. Pyrosequencing was performed using a 454 GS FLX titanium platform with sample-specific multiplex identifier sequence-adapted libraries for Lib-1 sequencing (Roche 454 Life Sciences, Branford, CT). Raw sff data files were first uncoupled into individual patient sample files using SFFFile tools (Roche). Low-quality reads and reads shorter than 90% of the expected amplicon lengths were removed.

The resultant data files were sequentially processed through implementation of the k-mer error correction (KEC) and empirical threshold algorithms as previously described using the parameters k = 25 and i = 3 (396, 397). The panel of temporally matched clonal sequences to the UDPS data was used to further identify and correct homopolymer errors (302, 396). Following this procedure, no erroneous sequences at a frequency > 0.1% were present in the homogeneous plasmid control sample. Consequently, all haplotypes present at a frequency > 0.1% in their respective sample were retained for downstream analysis.

2.6.2 1-step and k-step network construction.

To study the dynamics of intra-host quasispecies evolution, we created two networks for each patient (398). First, all unique haplotypes (318 bp) were aligned and the Hamming distance between each pair was calculated. Then connected components were built where each unique haplotype was represented by a node and two nodes were connected by an edge if the distance between them was one.

The 1-step network of most patients consisted of several connected components. To join them together, k-step networks were constructed as follows: iteratively for k = 2,3,…, until all pairs of haplotypes from different components with distance equal to k were found. They were linked by edges and the connected components were recalculated. These steps were repeated until a single component was formed. The resulting k-step network is equivalent to the union of all minimum spanning trees. The analysis and network visualization was performed with MATLAB R2014b (The MathWorks, Inc.) and Pajek (399).

2.6.3 Bioinformatics analyses.

MEGA6 was used to calculate Hamming distance, synonymous and nonsynonymous mutation rates (400). Phylogenetic trees using a general time-reversible model with gamma-distributed and invariant sites were drawn in MEGA6. Time ordered Shannon diversity (*H*) of 1-step networks was calculated using the formula:

$$H = -\sum_{i=1}^{N} p_i \times lnp_i$$

where p_i is the total frequency of haplotypes component *i* in the 1-step network and *N* is the number of connected components of the 1-step network (richness). Evenness (E_H) of the 1-step network was determined using the formula:

$E_H = H / \ln N$

Three patients were identified as containing mixed lineages. In each instance the components comprising the dominant lineage were analyzed separately from the minor lineages. Prior to calculation of Shannon diversity the total frequency of the dominant lineage components was normalized to 1 to account for the absence of the minor lineage.

Amino acid conservation plots were drawn using the Jalview program which is based on analysis of multiply aligned sequences (AMAS) to determine changes to the physio-chemical properties of the constituent amino acids (401, 402).

2.7 Methods in Chapter 8

We examined the serum of six patients with chronic HCV infection over a standardised interval of 16 weeks.

2.7.1 Separation of sample into Immunoglobulin enriched/ depleted fractions

Qproteome Albumin/IgG Depletion Kit (Qiagen, Venlo, Netherlands) protocol was modified for the separation of the sample into immunoglobulin enriched and immunoglobulin depleted fractions was performed as follows:

1. 5 ampliprep tubes were labelled alpha numerically in accordance with the subject under evaluation and the timing of the sample under investigation and this was followed by a number ranging from 1-5 – (e.g. for subject A at the time 16 weeks pre treatment the samples would be labelled A16-1, A16-2, A16-3, A16-4, A16-5).
2. The sample was thawed from the - 70°C freezer and mixed by gently pipetting.
3. 75µL phosphate buffering solution (PBS) and 25µL of the serum sample were transferred to a clean, labelled 1.5mL eppendorf tube and mixed by gently pipetting to produce the serum solution.
4. The Qproteome Depletion Spin Column was briefly centrifuged at 500g to remove the resin from the cap.
5. The screw cap was removed and the bottom broken from the spin column. The storage buffer was allowed to drain by gravity flow.
6. The spin column was equilibrated with 2 x 0.5mL aliquots of PBS and allowed to drain by gravity flow each time.
7. The spin column was centrifuged at 500g for 10 seconds with the luer plug in place.
8. 100µL serum solution (from step 3) was applied to the spin column, before the cap was secured and the column shaken vigorously until a viscous, homogenous suspension was produced.
9. The spin column was incubated on the end-over-end shaker for 5 minutes at room temperature.
10. The luer plug was removed and the column was transferred to the Ampliprep tube labelled A16-1.
11. The spin column cap was loosened (to avoid vacuum) ant the column was centrifuged at 500g for 10 seconds.
12. The spin column was transferred to the Ampliprep tube labelled A16-2 and washed with 2x 100µL PBS aliquots and centrifuged each time at 500g for 10 seconds.

13. The spin column was transferred to the Ampliprep tube A16-3 where step 12 was repeated.
14. Step 13 was repeated for A16-4 and A16-5.
15. 300µL lysis/binding buffer from the MagNa Pure LC Total Nucleic Acid Isolation Kit was added to both A15-3 and A16-4 and these were subsequently stored at -70°C.
16. A16-1, A16-2, and A16-5 were made up to 500µL with Lysis/Binding buffer. These samples contained the immunoglobulin free virus particles in sample A16-1 while the remaining two samples were included in further analysis to confirm that all immunoglobulin free virus was within the first sample.
17. Ampliprep was labelled A16-B
18. 200µL PBS buffer was added to 600µL lysis/binding buffer in a clean labelled 1.5mL eppendorf tube to generate a PBS/lysis/binding buffer solution.
19. The spin column (from step 14) was sealed with the luer cap and 200µL of the solution generated in step 28 was added and this was shaken vigorously and incubated on an end over end shaker for 5 minutes.
20. The spin column was transferred to the Ampliprep tube from step 17 (which had been labelled A16-B) and centrifuged for 10 seconds at 500g.
21. Step 19 and 20 were repeated.
22. The contents of A16-B were made up to 500µL with PBS/lysis/binding solution. This produced the immunoglobulin bound virus particles.

2.7.2 RNA extraction and nested E1/E2 PCR

These were performed as per Chapter 4.

2.7.3 Partitioned analysis of Quasispecies.

As per Chapter 4.

2.8 Methods in Chapter 9

2.8.1 Interferon gamma inducible protein 10 (IP 10) assay

IP 10 levels were measured using Luminex xMAP® bead based assay platform (Merck Millipore).

Serum samples were thawed vortexed for 1 minute and centrifuged for 5 minutes at 3,000 x g.

To pre-wet the plate, 150 µL wash buffer was used.

Pipette technique involved expression to the sides of the wells ensuring sure all fluid was expressed out of the pipette tips.

Luminex® colour code microspheres coated with specific IP 10 antibodies were incubated with the samples. Once captured by the bead, a biotinylated detection antibody is added and the reaction mixture is incubated with a reporter molecule which completes the reaction on the each bead.

The Luminex® reporter analyses the concentrations of each completed reaction within each well using LED based analysis. Quality controls with defined concentrations of the target molecule are included in parallel.

2.8.2 RT PCR

Reverse transcription of all sequences was performed as per Chapter 4.

2.8.3 Core nested PCR

Genotype 1

Primary PCR

Forward 5'-ATT GGG GGC GAC ACT CCA CCA T-3'

Reverse 5'-CGT AGG GGA CCA GTT CAT CAT CAT-3'

A 50 µl master mix was produced combining 5.0µl RT PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 3.0µl $MgSO_4$, 32.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 3 minutes, followed by 35 cycles of 94°C for 30 s, 64°C for 30 s, 72°C for 45 s and a final extension at 72°C for 7 minutes.

Secondary PCR

Forward 5'-CTT GTG GTA CTG CCT GAT AGG GTG C-3'

Reverse 5'-CCA RYT CAT CAT CAT RTC CCA NGC CA-3'

A 50 µl master mix was produced combining 5.0µl primary PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 38.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 s, 60°C for 45 s, 68°C for 2 min and a final extension at 68°C for 5 minutes. This yielded a product of 1,000 bp.

Genotype 3

Primary PCR

Forward 5'-CTT GTG GTA CTG CCT GAT AGG GTG C-3'

Reverse 5'-CCA RYT CAT CAT CAT RTC CCA NGC CA-3'

A 50 µl master mix was produced combining 5.0µl RT PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 38.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 s, 60°C for 45 s, 68°C for 2 min and a final extension at 68°C for 5 minutes.

Secondary PCR

Forward 5'-CTT GTG GTA CTG CCT GAT AGG GTG C-3'

Reverse 5'-CCA RYT CAT CAT CAT RTC CCA NGC CA-3'

A 50 µl master mix was produced combining 5.0µl primary PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 38.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 94°C for 30 s, 60°C for 45 s, 68°C for 1:30 min and a final extension at 68°C for 5 minutes. This yielded a product ~570bp.

2.8.4 Interferon Sensitivity Determining Region (ISDR) PCR

Genotype 1

Primary PCR

Forward 5'-CAG TGC TCA CTT CCA TGC TCA-3'

Reverse 5'-ACG GAT ATT TCC CTC TCA TCC-3'

A 50 µl master mix was produced combining 5.0µl RT PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 3.0µl $MgSO_4$, 32.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 95°C for 2 minutes, followed by 30 cycles of 95°C for 30 s, 55°C for 40 s, 72°C for 60 s and a final extension at 72°C for 7 minutes.

Secondary PCR

Forward 5'-ACC CCT CCC ACA TTA CAG CAG-3'

Reverse 5'-CCG AAG CGG ATC GAA AGA GTC CA-3'

A 50 µl master mix was produced combining 5.0µl primary PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 10.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 33.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 2 minutes, followed by 35 cycles of 95°C for 30 s, 55°C for 40 s, 72°C for 60 s and a final extension at 72°C for 7 minutes.

Genotype 3

Primary PCR

Forward 5'-TGC TGA GTT CTT TAC TGA-3'

Reverse 5'-GGT AAG GCG CAT CCA TGA A-3'

A 50 µl master mix was produced combining 5.0µl RT PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 10.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 33.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 5 minutes, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 45 s and a final extension at 72°C for 7 minutes.

Secondary PCR

Forward 5'-AGG GTG GAT GGG GTG AGA CTC AGT-3'

Reverse 5'-AGT CTG GCC TAG CCC AGA TAG GAA-3'

A 50 µl master mix was produced combining 5.0µl primary PCR product, 1.5µl forward primer, 1.5µl reverse primer, 1.0µl dNTP, 5.0µl 10x PCR buffer, 2.0µl $MgSO_4$, 29.5µl H_2O, and 0.5µl Pwo DNA polymerase. The reaction was performed in a thermocycler with a denaturation of 94°C for 5 minutes, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 45 s, and a final extension at 72°C for 7 minutes. This yielded a product ~933bp.

2.8.5 IL28 Sequencing

Samples were obtained from all study participants with informed consent and sequencing of the SNPs was outsourced.

2.9 Summary of Project Outputs

Subject	Amplicon	Prospective	Retrospective	Pyrosequencing	Core	ISDR	IL 28	IP 10
A	N	16-14-10-8-4-2-0	Y	Y	Y	Y	Y	Y
B	16-12-6-4-0	16-14-12-6-4-2-0	Y	Y	Y	Y	Y	Y
C	16-12-8-0	16-14-12-8-0	Y	Y	Y	Y		Y
D	16-12-8-4-0	16-14-12-8-6-4-2-0	Y	Y	Y	Y	Y	Y
E	N	16-14-12-8-4-0		Y	Y		Y	Y
F	16-12-8-4-0	16-14-12-8-6-4-2-0	Y	Y	Y	Y		Y
G	N	16-14-10-8-6-4-2-0	Y		Y	Y	Y	Y
H	N	16-14-12-10-8-6-4-2-0	Y	Y	Y		Y	Y
I	N	16-14-12-10-8-6-4-2-0	Y		Y	Y		Y
J	N	16-14-12-10-8-6-4-2-0	Y		Y	Y	Y	Y
K	16-12-8-4-0	14-12-10-8-6-4-2	Y		Y		Y	Y
L	16-8-4-0	16-14-10-8-6-4-2-0			Y		Y	Y
M	16-12-8-4-2-0	16-12-10-8-6-4-0	Y	Y	Y	Y		Y
N	16-2-0	16-2-0	Y		Y	Y	Y	Y
O	N	16-14-12-8-6-4-0	Y	Y	Y	Y	Y	Y
P	N	16-12-10-8-6-4-2-0			Y	Y		Y
Q	16-14-10-6-4-0	16-14-12-10-8-4-2-0	Y	Y	Y	Y		Y
R	N	16-14-12-10-8-6-4-0	Y		Y	Y		Y
S	N	16-14-12-8-6-4-0			Y	Y	Y	Y
T	16-12-8-4-0	16-14-12-8-4-2-0		Y	Y	Y	Y	Y
U	N	14-12-10-8-6-4-2-0	Y		Y	Y	Y	Y
V	N	14-12-10-8-6-4-2-0	Y		Y	Y	Y	Y
W	N	16-14-12-10-8-6-4-0	Y	Y	Y	Y	Y	Y

Table 1 – Overview of the results generated for each subject included in the project. Numbers correspond with the timing of the sample ore commencement of treatment.

Chapter 3

Published Hypothesis Paper

Hepatitis C quasispecies adaptation in the setting of a variable fidelity polymerase

Virus Adaptation and Treatment

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Hepatitis C quasispecies adaptation in the setting of a variable fidelity polymerase

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

This article was published in the following Dove Press journal:
Virus Adaptation and Treatment
23 July 2012
Number of times this article has been viewed

Daniel Schmidt-Martin[1]
Orla Crosbie[2]
Elizabeth Kenny-Walsh[2]
Liam J Fanning[1]

[1]Molecular Virology Diagnostic and Research Laboratory, Department of Medicine, University College Cork, Clinical Sciences Building, Cork University Hospital, [2]Department of Gastroenterology and Hepatology, Cork, Ireland

Abstract: Hepatitis C (HCV) is a virus characterized by an RNA-dependent RNA polymerase that lacks a proofreading mechanism and, as a result, generates a quasispecies. There is emerging evidence that this RNA-dependent RNA polymerase may in fact have variable fidelity. Here, we review the relevant concepts, including fitness landscapes, clonal interference, robustness, selection, adaptation, mutation rates, and their optimization, and provide a unique interpretation of a number of relevant theoretical models, evolving the theory of replicative homeostasis in light of their findings. We suggest that a variable fidelity polymerase can find its own optimal mutation rate, which is governed by the sequence itself and certain population dynamics. We propose that this concept can explain features of viral kinetics and clearance, both spontaneously and following treatment of chronic HCV. We point to evidence that supports this theory and explain how it refines replicative homeostasis and conclude by discussing particular areas of potential research that might augment our understanding of viral host interactions at an individual cellular level.

Keywords: fitness landscapes, adaptation, evolution, quasispecies, hepatitis C, replicative homeostasis

Introduction

Hepatitis C (HCV), a positive 9.2–9.6 kb RNA *Flavivirus*, was first identified by Choo and colleagues at Chiron in 1989 and is estimated to infect up to 3% people worldwide, equivalent to 120–170 million people.[1–3] Chronic HCV infection leads to the development of cirrhosis in 20% of cases after 20 years and is now the leading indication for orthoptic liver transplantation in the USA.

Low fidelity and the lack of proofreading ability of the HCV RNA-dependent RNA polymerase (RDRP) results in a population of closely related genomes or quasispecies.[4] Originally proposed by Eigen as a model for the study of the evolution of primitive organisms, the quasispecies concept has been applied to many bacteria and viruses including human immunodeficiency virus and HCV.[5] The gradual generation of point mutations results in the development of new variant species or "quasispecies" with slightly altered characteristics that then undergo selection. Within a given host, those quasispecies best adapted to the environment are most likely to survive and become dominant as a result of the principle of competitive exclusion.[6] The most prevalent quasispecies is the "master" sequence and other related quasispecies cluster around this in terms of their genetic distance. Mutations either undergo selection (positive, resulting in the selection of beneficial traits; negative, when a deleterious trait is removed). Alternatively, in the absence of selection, the gradual accumulation of

Correspondence: Daniel Schmidt-Martin
Molecular Virology Diagnostic and Research Laboratory, Department of Medicine, University College Cork, Clinical Sciences Building, Cork University Hospital, Cork, Ireland
Tel +353 21 490 1226
Fax +353 21 490 1227
Email d.schmidt@ucc.ie

© 2012 Schmidt-Martin et al, publisher and licensee Dove Medical Press Ltd. This is an Open Access article which permits unrestricted noncommercial use, provided the original work is properly cited.

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

neutral or near neutral mutations of insufficient magnitude to prompt selection results in an evolutionary process known as "genetic drift." Gradual adaptation to the host occurs as a result of these processes with the neutral theory of evolution predicting that genetic drift will be the predominant form of evolution.[7] For a mutation to provoke a change from genetic drift to natural selection it must breach what has been termed the "selection threshold."[8]

These processes are dependent on a number of factors including virion fitness, population size, clonal density, clonal interference, and mutation rates. Our understanding of HCV has grown exponentially as a result of both experimental results and mathematical modeling, which have facilitated better understanding of viral replication processes and, as a result, viral genomic selection, adaptation, and evolution. This review provides an up-to-date appraisal of these topics.

Fitness

Conceptually, each genome has an inherent fitness defined by a group of characteristics (ability to infect, ability to replicate, energy requirement), with each quasispecies competing for host resources (host cells, cellular machinery, etc). Within the population, each sequence competes for these host resources with the best adapted, or fittest characteristics, most likely to dominate. However, the transience of this domination is guaranteed by the mechanism by which it is generated; the almost inexorable emergence of fitter mutants demands continual evolution for survival, in a process called the "Red Queen Hypothesis."[9] A moderate increase in viral fitness of one quasispecies over another results in exponential proliferation of this new quasispecies, with likely extinction of its competitor sequences in what amounts to a zero-sum game.[10]

Fitness landscapes

Viral fitness can be described in the form of a fitness landscape, with mountains corresponding to areas of increased fitness surrounded by areas of diminishing fitness analogous to foothills (Figure 1). The accumulation of mutations allows the exploration of the sequence space and through this process the discovery of fitness gains that might displace the master sequence through competition. In the case of HCV, because the number of possible nucleotide combinations is so great ($4^{9.600}$), this landscape is only able to describe quasispecies diversity for short segments of the sequence. Not all mutations are viable, with these lethal mutations akin to cliffs in what is known as a "truncated fitness landscape." Finally, the combined interplay between individual quasispecies and the immune system results in a changing or dynamic truncated fitness landscape.

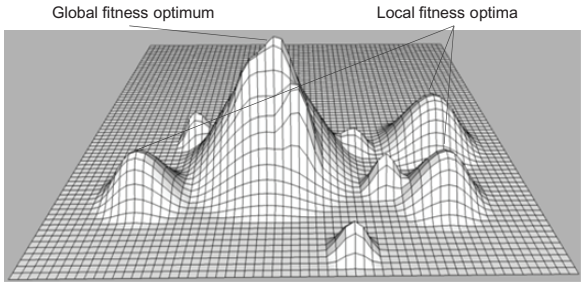


Figure 1 Schematic representation of the sequence space in the form of a fitness landscape.
Notes: The accumulation of mutation facilitates the exploration of the landscape. Adaptation results in the discovery of local fitness optima and potentially the global fitness optimum.

In this setting, the lack of a proofreading function is often looked upon as beneficial to HCV; adapted mutants, which are closely related to the parent virion and better able to evade the host's immune response, emerge and maintain chronic infection. However, in this model the proviso is that high mutation rates mean that beneficially adapted mutants are equally prone to deleterious mutations, which can potentially wipe out entire quasispecies. Muller's ratchet predicts that deleterious mutations are likely to "hitchhike" and be found in all future progeny, barring the unlikely event of a reciprocal mutation taking place.[11,12] Mitigating the effects of hitchhiking is the process of recombination, which can facilitate the removal of deleterious mutations by combining mutation-free segments and allow greater potential exploration of the sequence space by combining sequences with multiple mutations. It is this latter process that is thought to contribute significantly to the emergence of differing HCV genotypes and even taxa and species.[13,14]

Interference

Although a significant factor in determining the fate of a given quasispecies, competitive exclusion is not the sole determinant of evolutionary success. In large quasispecies populations, it has been shown that sequences with significant fitness superiority are not necessarily guaranteed to dominate a quasispecies due to a process known as "clonal interference."[15–17] In small populations, beneficial mutations of smaller increments are more likely to come to dominate as a result of selective sweep, while, in large populations (as are seen in established chronic HCV infections), a quasispecies with significant fitness benefit can be suppressed by the less-fit dominant quasispecies, unless it reaches a

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

critical threshold. This has the net effect of ensuring that, in chronic infection, the incremental increase in quasispecies fitness becomes larger in fitness gain but more infrequent in occurrence. Experimental evidence of clonal interference supporting this theory has been found in *Escherichia coli*, DNA viruses, HCV, and the RNA vesicular stomatitis virus (VSV).[16,18–21]

Defective interfering particles

Notwithstanding the extreme variability seen in the genetic sequence of RNA viruses, it must be remembered, however, that redundancy in the sequence is limited and that the proteins produced are small in number and, in most cases, essential in function. However, despite this lack of redundancy, subgenomic particles exist that can have significant effects on virus population dynamics.

Named "defective interfering particles" (DIPs) and identified in several virus species (including both DNA and RNA viruses), they may be important factors in the search for fitter quasispecies resistant to the effects of DIPs.[22–26] Unable to replicate in the absence of wild-type virus but able to infect new cells, they are thought to contribute to the oscillating nature of the viral load repeatedly seen in HCV infection. DIPs have also been proposed to interfere in the production of wild-type virus and modulate pathogen virulence and may themselves be potential antiviral agents.[27–30]

Stumpf and Zitzmann have proposed the reciprocality of DIPs; that is, that the particles are able to replicate but are unable to cause de novo infection of new cells due to the deletion of the structural section of the genome. The associated increase in replicative ability leads to competitive exclusion of viable virions and the gradual accumulation of defective intracellular viral RNA, meaning that continuous de novo infection of new cells is essential to viral survival.[31] Experimental evidence for this has remained elusive.

Studies focused on hepatocyte-derived HCV genomic sequences have not found evidence of these particles, though factors such as the duration of infection and use of limited numbers of clones (it is estimated that use of 20 clones will demonstrate most sequences present at a level of 10%) may go some way to explain this.[32–34] Indeed, the advent of next-generation sequencing may see the reemergence of this concept.

Robustness

The ability of a virion to tolerate mutations without phenotypic disruption, termed "robustness," is also likely to be important in maintaining or enhancing fitness. Characterized by a greater number of available neutral mutations, a high degree of robustness results in a smoother fitness landscape, in a theory described as "survival of the flattest."[35] Studies using digital models and subviral particles suggest that an organism with greater robustness may out compete and dominate less robust counterparts, particularly at times of high mutagenesis.[35–37]

The emergence of neutral mutation-rich organisms may however have significant implications for virion evolution. A recent study has demonstrated that a high proportion of neutral or near-neutral mutations may act as a barrier to evolution by natural selection, with genetic drift coming to dominate.[8] Studies evaluating HCV robustness are limited, but one network-based analysis of HCV polyprotein has demonstrated a high degree of robustness at many nucleotide positions, with relatively few positions vulnerable to phenotypically deleterious mutation.[38] Comparisons with other RNA viruses are challenging, as direct studies have not yet been published. One recent paper has estimated by site-directed mutagenesis that 40% of random mutations in VSV are lethal, which may suggest a lesser degree of robustness compared with HCV.[39]

Finally, it has been suggested that increased robustness may result in a reduced ability to adapt and that, in organisms that are required to survive in changing environments, the requirement for frequent adaptive change will limit tolerance of neutrality/robustness.[40] Indeed, the ability of an organism to respond to selective pressure and tolerate significant large-scale genetic evolution or evolvability has also been demonstrated as a selectable trait.[41]

Cooperative interaction

The concept of the "cooperative interaction" of the constituent mutants in exploring fitness maxima, so that the population ultimately achieves a mutation–selection equilibrium, distinguishes quasispecies theory from classical population genetics. When looked on in this light, it becomes apparent that successful quasispecies evolution is a population-wide phenomenon, so that fitness can be seen as an "ensemble property."[42] While evidence for this phenomenon is limited, studies of poliovirus have demonstrated that the pathogenesis of individual quasispecies is affected by cooperative interaction with other mutants in the quasispecies profile and that maintenance of a degree of heterogeneity is preferable for viral survival and maintenance of tissue tropism.[43] Indeed, the influence of cooperative reactions has been proposed as essential if mathematical models are to accurately generate the quasispecies patterns observed in vivo.[44]

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

Adaptation

"Adaptation" is the process whereby the quasispecies alters to become more suited to new or changing environments. The rate of adaptation of quasispecies appears to be governed by a number of factors: population size, mutation rate, adaptive quotient, and variability of the environment. In small populations, the size of the population limits the ability to explore the sequence space. As a result, adaptation occurs at a slower rate by means of stochastic genetic drift with episodic selective sweeps. This means that the population is more likely to be confined to local fitness peaks. In contrast, large populations are better able to expand throughout the sequence space and, as a result, adaptation is more deterministic, though the time taken for fixation of beneficial mutations is increased as a result of interference.[45,46]

Mutation rates

The effects of different mutation rates on a quasispecies within a truncated fitness landscape appear to follow three patterns: (1) low mutation rates result in a distribution around the master sequence and are more likely to become "trapped" in local fitness peaks, reducing the chances of complete exploration of the sequence space; (2) intermediate mutation rates result in wider exploration of the sequence space, with the emergence of variants further removed from the master sequence; and (3) those with high mutation rates produce an ever-increasing number of progeny with lethal mutations and, as a result, reach what has been called the "error threshold" – the point at which the quasispecies becomes unable to maintain sequence integrity. The coercion of viruses beyond this point into what has been called "error catastrophe" has been a major strategy in the development of antiviral therapies.[46–49]

Interestingly, adaptation is not maximal prior to reaching error threshold; rather, it behaves in a sine wave fashion (Figure 2). The mutation rate that results in optimal adaptation is remote from error threshold and adaptation decreases with increasing mutation rate as the ability to fix beneficial mutations decreases until error threshold is breached.[50,51] Using parameters present in quasispecies, Orr has found that optimal mutation rates for adaptation are governed by the strength of selection against deleterious mutations (ie, more truncated landscapes have a lower optimal adaptive mutation rate).[50] This work was based on the assumption that the selective power against deleterious mutations was at all times greater that the selective power for beneficial mutations; however, this may not, in fact, be the case.

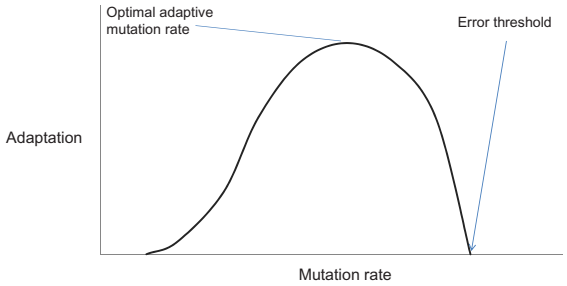


Figure 2 The interaction between mutation rates and the rate of adaptation.
Notes: Beyond the optimal adaptive mutation rate, increasing mutation rates result in an ever-diminishing rate of adaptation until the error threshold is reached.

Adaptive quotient

Johnson and Barton advanced this theory by describing a matrix that can predict a sequence-specific optimal mutation rate depending on whether the surrounding fitness landscape is dominated by deleterious or beneficial mutations and the selective power of these relative to each other.[52] According to this model, the existence of many potential beneficial mutations will promote the emergence of a higher mutation rate and vice versa. In the setting of a bottleneck event (rapid reduction in quasispecies – as occurs at transmission of HCV) the organism can be seen to be less adapted to the new host and, as a result, the ratio of beneficial:deleterious mutations is also likely to change and will probably be reflected in the rate of mutation.[53]

Variability of environment

In static environments, the exploration of the sequence space with fixation of beneficial mutations that pass the selection threshold and outlast clonal interference becomes exhausted once the quasispecies reaches the mutation–selection equilibrium. At this stage, all fitness optima have been explored. As this occurs, the fitness gains that were initially large, diminish toward nil.[54] With many microorganisms, however, the emergence of new environments, either as a result of transmission of infection or the development of immune responses, results in a dynamic fitness landscape that serves to replenish the potential for adaptive change.

Furthermore, the ruggedness of these landscapes can themselves affect the rate of adaptation. Clune et al demonstrated using computer models that digital organisms fail to optimize mutation rates and tend to settle at a mutation rate below this.[55] Evidence for this has been described in DNA bacteriophages, where the imposition of a fourfold increase in mutation rate actually conferred fitness gain.[55,56] Clune et al argued that, while adaptation occurs over long periods (many generations), selection acts quickly and this phenomenon

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

may be an effort by the virus to mitigate the potential for emergence of deleterious/lethal mutations. Expanding on this initial finding, Clune et al demonstrated that the observed mutation rate is dependent on the ruggedness of the fitness landscape with more rugged landscapes favoring an even lower mutation rate.[55]

The fidelity spectrum

There is a growing body of literature indicating that mutation rates are not constant and may be selectable. Mutation rates have been shown to increase at times of stress in many bacteria and lethal mutagenesis has long been suggested as a potential treatment strategy in viral infections.[57] The beneficial effects of increased mutational rates, in addition to how they may be associated with increased replicative capacity, have also been demonstrated in bacteriophage populations.[58] Furthermore, adaptive change in the mutation rate in response to medications has been shown to confer drug resistance and sustain chronic infection in the case of human immunodeficiency virus type 1.[59] Several mechanisms governing how transient increases in mutation rate can be generated and suitably regulated have been suggested, including environmental and heritable factors.[60]

Evidence of variable RDRP mutation rates in HCV

In HCV, the estimated mutation rate is $1 \times 10^{-4 \text{ to } 5}$/base.[61–63] The estimation of the error threshold of HCV RDRP is $10^{-2 \text{ to } 3}$ (mutations per base), which leaves scope for a ten- to hundredfold change in baseline RDRP fidelity before the error threshold is reached, with the optimal adaptive mutation rate likely to be found within this range. Lethal mutagenesis has formed one of the theories of the mechanism of action of ribavirin, as it has been shown to induce lethal mutagenesis in poliovirus and foot-and-mouth disease, but the results in HCV have been variable.[62,64,65] Ribavirin-resistant mutations have been described, both in vitro and in vivo, in two HCV nonstructural proteins (NS5A and NS5B) including the RDRP. It has been suggested that the NS5A mutation may indicate that this protein may interact with the RDRP to modulate polymerase fidelity.[62,66,67] Indeed, the idea that RDRP fidelity may be controlled remotely is not novel to HCV.[68]

Recently, mutant RDRPs conferring ribavirin resistance by means of increased fidelity have been described in both foot-and-mouth disease virus and poliovirus. We feel that similar mutants are likely to exist in the case of HCV and that their emergence during ribavirin therapy would explain the heterogeneity of the effect on mutation rates seen in these studies. Furthermore, the sampling intervals may have been such that transient increases in mutation rates may have been missed.

A mechanism by which viruses might self-regulate replication fidelity has been proposed by Sallie in his theory of replicative homeostasis (RH).[69–71] Sallie argued that HCV viral kinetics behave in such a way to suggest autoregulation of virion production through a homeostatic mechanism that modulates RDRP fidelity/processivity (which he proposed are inversely proportional). RH predicts that excess wild-type protein will prompt a decrease in fidelity and a resultant increase in mutation and vice versa. The idea that mutation rates may be dependent on polymerization rates was first proposed in the kinetic proofreading hypothesis, in which a delay in the rate of polymerization results in increased polymerase fidelity; experimental support for this has been demonstrated in the case of VSV.[11,72] While we are of the opinion that Sallie's theory has significant merit, we feel that the theory of RH could be further adapted to more accurately describe the behavior of HCV.

A framework for the action of a variable fidelity polymerase

On the basis of the position of the quasispecies within the fitness landscape, we propose that a framework exists for the selection of phase-specific mutation rates (Figure 3). The RDRP acts along a fidelity spectrum with optimal mutation rates that are largely dependent on population size, capacity for adaption (adaptive quotient), and variability of

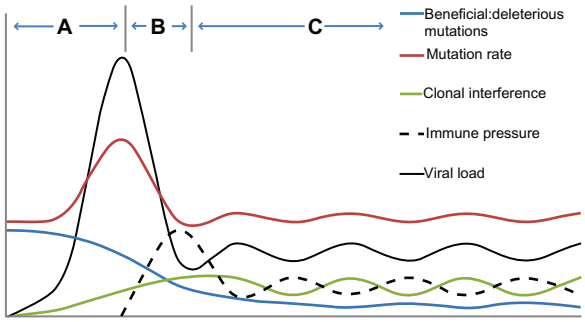


Figure 3 Phase diagram demonstrating the proposed behavior of hepatitis C (HCV) variable fidelity polymerase. (**A**) New infection is characterized by an increase in the ratio of beneficial:deleterious mutations. HCV polymerases with increased mutation rates are selected, promoting exploration of the sequence space, which results in viral load spike. (**B**) Once the sequence space is explored, the ratio of beneficial:deleterious mutations decreases and the polymerase mutation rate returns to baseline. As the quasispecies expands, clonal interference emerges. The advent of host immune response, in combination with reduced mutation, is associated with a marked reduction in viral load. (**C**) Immune-mediated dynamic changes in the fitness landscape result in oscillation of clonal interference, viral load, and polymerase mutation rates.

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

the environment.[73,74] We propose that the optimal mutation rate selected for could be predicted by the position of the sequence within a framework similar to that proposed by Johnson and Barton.[52] The exploration of this fidelity spectrum is likely to be initially stochastic, as it is reliant on the generation of promutator mutations and evidence for similar processes can be seen in *Drosophila* populations.[52,75]

Following the bottleneck of transmission, unencumbered by clonal interference, and with an increased probability of beneficial mutations, we suggest that a form of density-dependent selection, similar to those that have been described in foot-and-mouth disease virus, *E. coli*, and *Drosophila*, will result in the emergence of an increased mutation rate.[75–77] This latter occurrence would be characterized by quicker adaptation, could correspond to the intermediate fidelity phase as described by Saakian et al, and could be likened to the episodes of stress which have also been shown to result in increased mutation rates in *E. coli*.[47,78] As the relative proportion of beneficial to deleterious mutations is increased in small nonadapted populations, the emergence of an increased mutation rate is favored. Initial infection with a finite number of variants will gradually explore local fitness maxima by stochastic means until the population becomes sufficient for deterministic exploration as the capacity to generate all possible mutants is achieved.[46] With population expansion, increasing clonal interference, and viral adaptation, the same process will select a less-productive polymerase with increased fidelity that has the added potential bonus of being immunologically stealthy by means of viral-load reduction. This period of selection may result in the reduction in viral load often seen in acute HCV.[70]

Furthermore, increased fidelity will inevitably mean that antigenic thresholds will be intermittently breached, resulting in activation of the adaptive immune response. As the exploration of both sequence space and what we refer to as the "fidelity spectrum" is stochastic, it is to some degree dependent on chance, but the near certainty of successful exploration has been in built into the quasispecies characteristics of the virus. Conversely, the certainty of failure, in some cases, to either optimize fidelity or even find infidelity sufficient to evade the immune response, provides the tantalizing prospect of explaining the process by which infection is cleared in 15%–25% of patients.[79,80] The emergence of a population selection–mutation equilibrium will tend toward a lower mutation rate, as the genetic distance to the nearest beneficial mutations is likely to become larger due to this adaptation. In summary, at times of stress, the polymerase and its inherent mutation rate becomes the unit of selection, while, at other times, it is the genomic properties and their cooperative/competitive interactions that become the traits selected for or against.

In our model, similar to that of RH, the selection of particular sequences for removal by the immune system will merely result in the generation of new quasispecies to match the new fitness landscape, while also facilitating long-term stability of quasispecies in the absence of variations in effective immune pressure. This model also has the capacity to explain HCV clearance in the absence of seroconversion, as it allows for the attainment of error catastrophe without the need for immune response. Additionally, our proposed mechanism of action along a fidelity spectrum more coherently explains why the emergence of a single dominant quasispecies in the treatment of HCV infection and a low rate of quasispecies evolution are more likely to result in clearance as opposed to the generation of new quasispecies, as Sallie's model would suggest.[69–71]

In proposing this model, we must acknowledge that one of the major obstacles to clarifying the interaction between quasispecies theory and experimental results in HCV is the phenomenon of founder effect at the level of the individual cell. The prevention of superinfection, in theory, means that the apparatus of the cell is at the mercy of this sole founder and that competition is prevented, promoting the preservation of the status quo. This, coupled with evidence demonstrating the prevention of infection of neighboring cells via the apical cell membrane and the facilitation of virion transfer to these neighbors via tight junctions, is equivalent to dynasty building – that is, clonal expansion. Accounting for these factors in evolutionary models is challenging, particularly when little is known of the incidence of superinfection in the context of fitter "pilgrim" virions, which may facilitate the conversion of the quasispecies to new fitness optima.

Finally, we would like to note one conundrum reconciling the current theories of optimal mutation rates and the suggestion that organisms adapt toward neutral networks. Under the survival of the flattest hypothesis, the emergence of such fitness landscapes results in a reduction in the ruggedness of the fitness landscape. As the number and selective power of potentially deleterious mutations are reduced, we should see closer optimization of mutation rates to maximize adaptation. However, little evidence has been produced to favor this and, conversely, the mutation rate in *E. Coli*, which has a 90% tolerance of mutations, has a mutation rate far less than that of RNA viruses, which have a lethal mutation rate of 21%–40%.[39,81]

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

Conclusion

Adaptive evolution is slave to both genetic drift and natural selection, with the emergence of more neutral flatter fitness landscapes favoring the former. Following a bottleneck, the exploration of the sequence space is stochastic, with the transition to deterministic exploration dependent on the population size and the development of clonal interference. Mutation rates often fail to optimize adaptation and this may be an effort to mitigate the relative strength of lethal mutations when compared with the relative and often-marginal benefit of beneficial mutations – particularly, in well-already-adapted species. Mutation rates are not constant and, in low population sizes, increased mutation rates may be selected for to enhance the rate of adaptation. Several potential mechanisms for regulating mutation rates to ensure that these increases are transient have been proposed. HCV demonstrates characteristics consistent with a population density-mediated selection of mutation rates.

Acknowledgment

The authors would like to acknowledge the support of Daniel Schmidt-Martin from Molecular Medicine Ireland.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*. 1989;244(4902):359–362.
2. Alter MJ. Epidemiology of hepatitis C virus infection. *World J Gastroenterol*. 2007;13(17):2436–2441.
3. Pockros PJ. Developments in the treatment of chronic hepatitis C. *Expert Opin Investig Drugs*. 2002;11(4):515–528.
4. Steinhauer DA, Domingo E, Holland JJ. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene*. 1992;122(2):281–288.
5. Eigen M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*. 1971;58(10):465–523.
6. Gause GF. *The Struggle for Existence.* New York, Dover; 1971.
7. Kimura M. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press; 1983.
8. Nelson CW, Sanford JC. The effects of low-impact mutations in digital organisms. *Theor Biol Med Model*. 2011;8:9.
9. Van Valen L. A new evolutionary law. *Evol Theory*. 1973;1(1):1–30.
10. Solé RV, Ferrer R, González-García I, Quer J, Domingo E. Red queen dynamics, competition and critical points in a model of RNA virus quasispecies. *J Theor Biol*. 1999;198(1):47–59.
11. Furió V, Moya A, Sanjuán R. The cost of replication fidelity in an RNA virus. *Proc Nat Acad Sci U S A*. 2005;102(29):10233–10237.
12. Muller HJ. The relation of recombination to mutational advance. *Mutat Res*. 1964;106:2–9.
13. Worobey M, Holmes EC. Evolutionary aspects of recombination in RNA viruses. *J Gen Virol*. 1999;80(Pt 10):2535–2543.
14. Koonin EV, Dolja VV. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol*. 1993;28(5):375–430.
15. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. *Genetica*. 1998;102–103(1–6):127–144.
16. Miralles R, Gerrish PJ, Moya A, Elena SF. Clonal interference and the evolution of RNA viruses. *Science*. 1999;285(5434):1745–1747.
17. de la Torre JC, Holland JJ. RNA virus quasispecies populations can suppress vastly superior mutant progeny. *J Virol*. 1990;64(12):6278–6281.
18. Miralles R, Moya A, Elena SF. Diminishing returns of population size in the rate of RNA virus adaptation. *J Virol*. 2000;74(8):3566–3571.
19. Imhof M, Schlötterer C. Fitness effects of advantageous mutations in evolving Escherichia coli populations. *Proc Nat Acad Sci U S A*. 2001;98(3):1113–1117.
20. Miller CR, Joyce P, Wichman HA. Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics*. 2011;187(1):185–202.
21. Verbinnen T, Van Marck H, Vandenbroucke I, et al. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. *J Virol*. 2010;84(21):11124–11133.
22. Bantel-Schaal U, zur Hausen H. Characterization of the DNA of a defective human parvovirus isolated from a genital site. *Virology*. 1984; 134(1):52–63.
23. Laughlin CA, Myers MW, Risin DL, Carter BJ. Defective-interfering particles of the human parvovirus adeno-associated virus. *Virology*. 1979;94(1):162–174.
24. Palma EL, Huang A. Cyclic production of vesicular stomatitis virus caused by defective interfering particles. *J Infect Dis*. 1974;129(4):402–410.
25. Prince AM, Huima-Byron T, Parker TS, Levine DM. Visualization of hepatitis C virions and putative defective interfering particles isolated from low density lipoproteins. *J Viral Hepat*. 1996;3(1):11–17.
26. Horikami SM, Curran J, Kolakofsky D, Moyer SA. Complexes of Sendai virus NP-P and PL proteins are required for defective interfering particle genome replication in vitro. *J Virol*. 1992;66(8):4901–4908.
27. Perrault J. Origin and replication of defective interfering particles. *Curr Top Microbiol Immunol*. 1981;93:151–207.
28. Kirkwood TB, Bangham CR. Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. *Proc Nat Acad Sci U S A*. 1994;91(18):8685–8689.
29. Marriott AC, Dimmock NJ. Defective interfering viruses and their potential as antiviral agents. *Rev Med Virol*. 2010;20(1):51–62.
30. Rao DD, Huang AS. Interference among defective interfering particles of vesicular stomatitis virus. *J Virol*. 1982;41(1):210–221.
31. Stumpf MP, Zitzmann N. RNA replication kinetics, genetic polymorphism and selection in the case of the hepatitis C virus. *Proc Biol Sci*. 2001;268(1480):1993–1999.
32. Navas S, Martín J, Quiroga JA, Castillo I, Carreño V. Genetic diversity and tissue compartmentalization of the hepatitis C virus genome in blood mononuclear cells, liver, and serum from chronic hepatitis C patients. *J Virol*. 1998;72(2):1640–1646.
33. Gretch DR, Polyak SJ. The quasispecies nature of Hepatitis C Virus: research methods and biological implications. In: Groupe Français d'études molécularies des hépatites (GEMHEP), *Hepatitis C Virus: 2; Genetic Heterogeneity and Viral Load*. Paris: John Libbey Eurotext; 1997:57–69. [English].
34. Cabot B, Martell M, Esteban JI, et al. Nucleotide and amino acid complexity of hepatitis C virus quasispecies in serum and liver. *J Virol*. 2000;74(2):805–811.
35. Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*. 2001;412(6844):331–333.
36. Wilke CO. Adaptive evolution on neutral networks. *Bull Math Biol*. 2001;63(4):715–730.
37. Codoñer FM, Darós JA, Solé RV, Elena SF. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog*. 2006;2(12):e136.
38. Campo D, Dimitrova Z, Mitchell R, Lara J, Khudyakov Y. Coordinated evolution of the hepatitis C virus. *Proc Nat Acad Sci U S A*. 2008; 105(28):9685–9690.

Virus Adaptation and Treatment downloaded from https://www.dovepress.com/ by 143.239.102.28 on 01-Sep-2017
For personal use only.

39. Sanjuán R, Moya A, Elena SF. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Nat Acad Sci U S A*. 2004;101(22):8396–8401.
40. Aguirre J, Lázaro E, Manrubia SC. A trade-off between neutrality and adaptability limits the optimization of viral quasispecies. *J Theor Biol*. 2009;261(1):148–155.
41. Earl DJ, Deem MW. Evolvability is a selectable trait. *Proc Nat Acad Sci U S A*. 2004;101(32):11531–11536.
42. Schuster P, Swetina J. Stationary mutant distributions and evolutionary optimization. *Bull Math Biol*. 1988;50(6):635–660.
43. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*. 2005;439(7074):344–348.
44. Arbiza J, Mirazo S, Fort H. Viral quasispecies profiles as the result of the interplay of competition and cooperation. *BMC Evol Biol*. 2010;10:137.
45. de Visser JA, Rozen DE. Limits to adaptation in asexual populations. *J Evol Biol*. 2005;18(4):779–788.
46. Jain K, Krug J. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics*. 2007;175(3):1275–1288.
47. Saakian DB, Biebricher CK, Hu CK. Phase diagram for the Eigen quasispecies theory with a truncated fitness landscape. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2009;79(4 Pt 1):041905.
48. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Nat Acad Sci U S A*. 2001;98(12): 6895–6900.
49. Biebricher CK, Eigen M. The error threshold. *Virus Res*. 2005;107(2): 117–127.
50. Orr HA. The rate of adaptation in asexuals. *Genetics*. 2000;155(2): 961–968.
51. Stich M, Briones C, Manrubia SC. Collective properties of evolving molecular quasispecies. *BMC Evol Biol*. 2007;7:110.
52. Johnson T, Barton NH. The effect of deleterious alleles on adaptation in asexual populations. *Genetics*. 2002;162(1):395–411.
53. Arjan JA, de Visser M, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. Diminishing returns from mutation supply rate in asexual populations. *Science*. 1999;283(5400):404–406.
54. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*. 2003;4(6): 457–469.
55. Clune J, Misevic D, Ofria C, Lenski RE, Elena SF, Sanjuán R. Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. *PLoS Comput Biol*. 2008;4(9):e1000187.
56. Springman R, Keller T, Molineux IJ, Bull JJ. Evolution at a high imposed mutation rate: adaptation obscures the load in phage T7. *Genetics*. 2010;184(1):221–232.
57. Foster PL. Adaptive mutation: implications for evolution. *Bioessays*. 2000;22(12):1067–1074.
58. Cases-González C, Arribas M, Domingo E, Lázaro E. Beneficial effects of population bottlenecks in an RNA virus evolving at increased error rate. *J Mol Biol*. 2008;384(5):1120–1129.
59. Wainberg MA, Drosopoulos WC, Salomon H, et al. Enhanced fidelity of 3TC-selected mutant HIV-1 reverse transcriptase. *Science*. 1996;271(5253):1282–1285.
60. Metzgar D, Wills C. Evidence for the adaptive evolution of mutation rates. *Cell*. 2000;101(6):581–584.
61. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics*. 1998;148(4):1667–1686.
62. Cuevas JM, González-Candelas F, Moya A, Sanjuán R. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. *J Virol*. 2009;83(11):5760–5764.
63. Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proc Nat Acad Sci U S A*. 1999;96(24):13910–13913.
64. Vignuzzi M, Stone JK, Andino R. Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. *Virus Res*. 2005;107(2):173–181.
65. Airaksinen A, Pariente N, Menéndez-Arias L, Domingo E. Curing of foot-and-mouth disease virus from persistently infected cells by ribavirin involves enhanced mutagenesis. *Virology*. 2003;311(2):339–349.
66. Pfeiffer JK, Kirkegaard K. Ribavirin resistance in hepatitis C virus replicon-containing cell lines conferred by changes in the cell line or mutations in the replicon RNA. *J Virol*. 2005;79(4):2346–2355.
67. Young KC, Lindsay KL, Lee KJ, et al. Identification of a ribavirin-resistant NS5B mutation of hepatitis C virus during ribavirin monotherapy. *Hepatology*. 2003;38(4):869–878.
68. Arnold JJ, Vignuzzi M, Stone JK, Andino R, Cameron CE. Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J Biol Chem*. 2005;280(27):25706–25716.
69. Sallie R. Replicative homeostasis III: implications for antiviral therapy and mechanisms of response and non-response. *Virol J*. 2007;4:29.
70. Sallie R. Replicative homeostasis: a fundamental mechanism mediating selective viral replication and escape mutation. *Virol J*. 2005;2:10.
71. Sallie R. Replicative homeostasis II: influence of polymerase fidelity on RNA virus quasispecies biology: implications for immune recognition, viral autoimmunity and other "virus receptor" diseases. *Virol J*. 2005;2:70.
72. Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Nat Acad Sci U S A*. 1974;71(10):4135–4139.
73. Simon B. A stochastic model of evolutionary dynamics with deterministic large-population asymptotics. *J Theor Biol*. 2008;254(4):719–730.
74. Bukh J, Pietschmann T, Lohmann V, et al. Mutations that permit efficient replication of hepatitis C virus RNA in Huh-7 cells prevent productive replication in chimpanzees. *Proc Nat Acad Sci U S A*. 2002;99(22):14416–14421.
75. Mueller LD, Ayala FJ. Trade-off between r-selection and K-selection in Drosophila populations. *Proc Nat Acad Sci U S A*. 1981;78(2): 1303–1305.
76. Giraud A, Matic I, Tenaillon O, et al. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science*. 2001;291(5513):2606–2608.
77. Sevilla N, Ruiz-Jarabo CM, Gómez-Mariano G, Baranowski E, Domingo E. An RNA virus can adapt to the multiplicity of infection. *J Gen Virol*. 1998;79(Pt 12):2971–2980.
78. Rosenberg SM, Thulin C, Harris RS. Transient and heritable mutators in adaptive evolution in the lab and in nature. *Genetics*. 1998;148(4):1559–1566.
79. Alter HJ, Conry-Cantilena C, Melpolder J, et al. Hepatitis C in asymptomatic blood donors. *Hepatology*. 1997;26(3 Suppl 1):29S–33S.
80. Alter HJ, Seeff LB. Recovery, persistence, and sequelae in hepatitis C virus infection: a perspective on long-term outcome. *Semin Liver Dis*. 2000;20(1):17–35.
81. Sanjuán R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc Lond B Biol Sci*. 2010;365(1548):1975–1982.

Virus Adaptation and Treatment Dovepress

Publish your work in this journal

Virus Adaptation and Treatment is an international, peer-reviewed open access journal focusing on the study of virology, viral adaptation and the development and use of antiviral drugs and vaccines to achieve improved outcomes in infection control and treatment. The journal welcomes original research, basic science, clinical & epidemiological studies, reviews & evaluations, expert opinion and commentary, case reports and extended reports. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/virus-adaptation-and-treatment-journal

Chapter 4

Published Study

Intensive temporal mapping of Hepatitis C hypervariable region 1 quasispecies provides novel insights into HCV evolution in chronic infection.

Schmidt-Martin, D., Crosbie, O., Kenny-Walsh, E. and Fanning, L. J. (2015) 'Intensive temporal mapping of hepatitis C hypervariable region 1 quasispecies provides novel insights into hepatitis C virus evolution in chronic infection', *Journal of General Virology,* 96(8), pp. 2145-2156. DOI: http://dx.doi.org/10.1099/vir.0.000149

Chapter 5

Does amplicon sequencing accurately reflect the underlying HVR1 quasispecies?

5.1. Background

HCV quasispecies are thought to enable the virus to chronically infect human hosts, as the virus mutant spectra facilitates the evasion of the host immune system(403). Hypervariable Region 1 which is found at the N terminal section of E2 has been extensively studied and is thought important in ongoing immune evasion by means of its malleability(313). Study into the characteristics of HCV QS among patients who underwent virus treatment with dual therapy comprising pegylated interferon and ribavirin have identified both HVR1 QS complexity and diversity as potential predictors of treatment success(344, 350, 354). We explored how accurately the PCR amplicon of a nested HVR1 PCR reflects the underlying quasispecies. In order to do this we cloned and amplified nested PCR products of HVR1 using a high fidelity DNA polymerase (*Pwo* DNA polymerase). In parallel with this process we sequenced the PCR product of 10 of the subjects in our study using *Taq* polymerase.

5.2. Methods

Serum was collected from 10 individuals as per the methods and guidelines outlined previously and the serum spun, collected and stored at -80°C within 6 hours of collection.

For a detailed description of methods used see:

Chapter 2 section 2.4

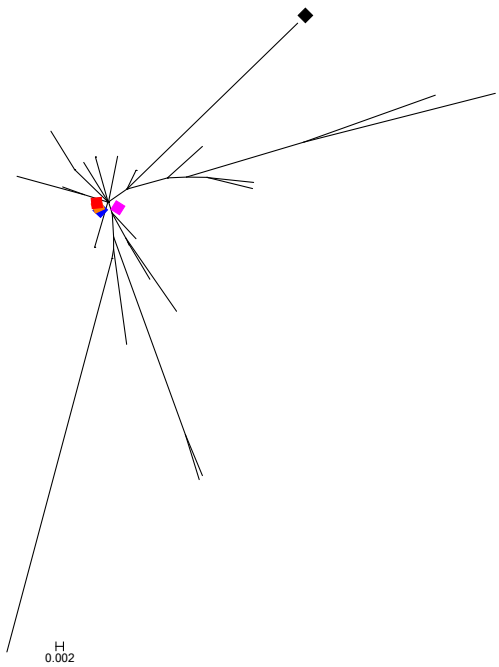
5.3. Results

The results for each individual are presented separately. The subjects are designated alphabetically and these correspond with those from chapter 4.

The amplicons generated are presented on phylogenetic trees generated by combining the amplicon sequences with all unique nucleotide HVR1 sequences from the prospective cloning project. Phylogenetic trees were constructed using maximum composite likelihood methods with a gamma distribution and each tree was boot strapped 10,000 times for the purposes of tree optimisation.

The phylogenetic tree is followed by multiple sequence alignments containing the amplicon sequence with the corresponding unique HVR1 clones for the same sample.

5.3.1 Subject B



5.3.1 Fig 1.

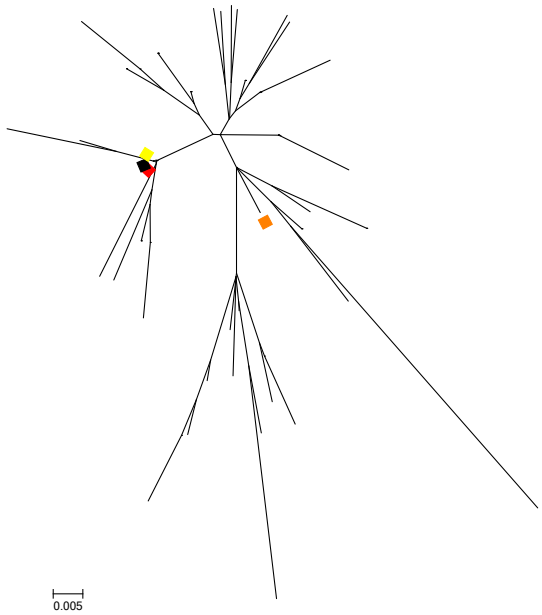
Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = week 16, Red = Week 12, Blue = Week 6, Pink = Week 4, Orange = Week 0. All sequences with the exception of the sample at week 16, which differs at numerous amoino acid positions, are identical at nucleotide level and therefore the square labels overlap on the phylogenetic tree (see 5.3.1 Fig 2). This suggests initial change followed by stasis of HVR1 in the study period.

```
                        1        10         20      27
                        |--------+---------+------|
          110616_IF-1   ATYTTGGATGHNTFRLTSLFDSGPQQK
          110660-16-2   ATYTTGGAQAYNTFRLTSLFDSGPQQK
110660-16-1310660-16    ATYTTGGAQAYNTFRLTSLFDSGPQQK
         110660-16-12   ATYTTGGAQAYHAFRLTSLFDSGPQQK
         110660-16-20   ATYATGGSDAYDTFRLTSLFDSGAQQK
          110660-16-5   ATYTTGGSQAHTTSRFTSFFDLGPQQK
            Consensus   ATYtTGGsqay.tfRlTSlFDsGpQQK


                        1        10         20      27
                        |--------+---------+------|
1106-1220_M13uni_--_    ATYATGGTQAYNTFRLASLFDSGPPQK
1106-1211_M13uni_--_    ATYATGGTQAYNTFRLVSLFDSGPPQK
1106-123_M13uni_--_5    ATYTTGGAQAYNTFRLTSLFDSGPQQK
 1106-121_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
          110612_IF-1   ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1219_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
              1106-122  ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1213_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-126_M13uni_--_9    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1210_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1214_M13uni_--_    ATYTTGGAQAYNAFRLTSLFDSGPQQK
            Consensus   ATYtTGGaQAYNtFRLtSLFDSGPqQK


                        1        10         20      27
                        |--------+---------+------|
  1106617_M13uni106-    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-6-4_M13uni_--_2    ATYTTGGAQAYNTFRLTSLFDSGPQQK
           11066_IF-1   ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-6-5_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-6-6_M13uni_--_6    ATYTTGGAQAYNTFRLTSLFDSGPQQR
        110666_M13uni   ATYTTGGAQAYNAFRLTSLFDSGPQQK
            Consensus   ATYTTGGAQAYNtFRLTSLFDSGPQQk


                        1        10         20      27
                        |--------+---------+------|
             8_M13uni   ATYTTGGAQAYNTFRLTSLFDSGPQQK
             3_M13uni   ATYTTGGAQAYNTFRLTSLFDSGPQQK
           11064_IF-1   ATYTTGGAQAYNTFRLTSLFDSGPQQK
      1160418_M13uni    ATYTTGGAQAYNTFRLTSLFDSGPQQK
            11_M13uni   ATYTTGGAQAYNTFRLTSLFDPGPQQK
            Consensus   ATYTTGGAQAYNTFRLTSLFDsGPQQK


                        1        10         20      27
                        |--------+---------+------|
1106-0-1_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-0-8_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
           11060_IF-1   ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-0-14_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDPGPQQK
            Consensus   ATYTTGGAQAYNTFRLTSLFDsGPQQK
```

5.3.1 Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 6, 4, 0. The amplicon sequence closely corresponds with the cloning data in all samples except week 16 where there is a 3 amino acid difference between the amplicon sequence and the nearest clone at positions 9-11.

5.3.2. Subject C



5.3.2. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = week 16, Red = Week 12, Yellow = Week 8, Orange = Week 0. The nucleotide sequences from Weeks 16, 12, and 8 are identical and the labels therefore overlap. The HVR1 change which results in the Week 0 sequence is characterised by a single amino acid substitution (see 5.3.2 Fig2.) which indicates minimal overall change in HVR1 during the entire study period.

Although the tree has multiple apparent clades, analysis on the amino acid sequences identifies that this reflects multiple synonymous nucleotide substitutions and the overall genetic distances is small between all sequences.

```
                        1       10        20      27
                        |--------+---------+------|
          131216_IF-1   ETHITGGTQARTTRGFANLFSPGPSQN
1312-1611_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-1613_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-1618_M13uni_--_    ETHITGGAQARTTRGFANLFSPGPSQN
            1312-1619   ETHITGGAQARTTRGFANLFSPGPSQN
1312-16-1_312-16-2_M    ETHITGGAQARTTRGFANLFSPGPSQN
            Consensus   ETHITGGtQARTTRGFANLFSPGPSQN
```

```
                        1        10         20      27
                        |--------+---------+------|
1312-12-1_M13uni312-    ETHITGGTQARTTRGFANLFSPGPSQN
1312-12-2_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
          131212_IF-1   ETHITGGTQARTTRGFANLFSPGPSQN
1312-1216_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-1211_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-124_M13uni_--_1    ETHITGGTQARTTRGFANLFSPGPSQN
            Consensus   ETHITGGTQARTTRGFANLFSPGPSQN
```

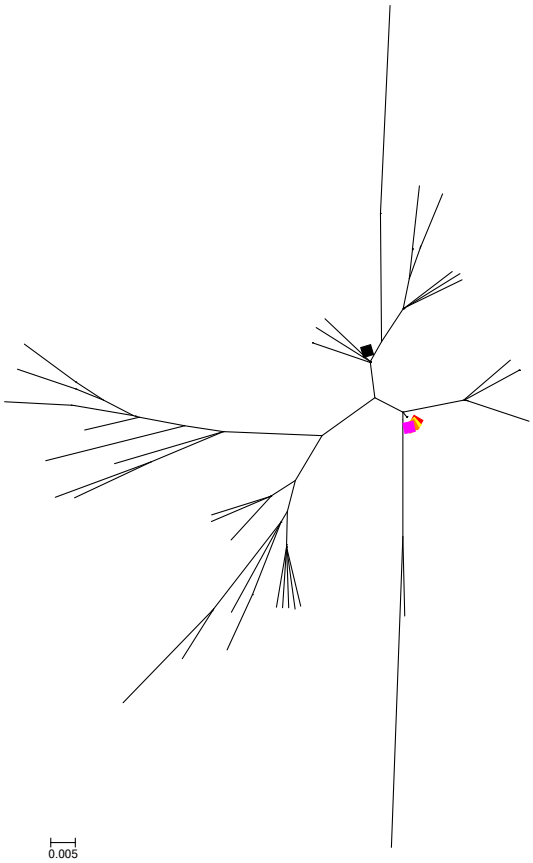
```
                        1        10         20      27
                        |--------+---------+------|
          1312-8-21_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-12_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
          13128_IF-1    ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-8_M13uni_--_1    ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-20_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-2_M13uni_--_1    ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-19_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-10_M13uni_--_    ETHITGGAQARTTRGFANLFSPGPSQN
1312-8-4_M13uni_--_2    ETHITGGAQARTTRGFANLFSPGPSQN
1312-8-5_M13uni_--_1    ETHITGGAQARTTRGFANLFSPGPSQN
            Consensus   ETHITGGtQARTTRGFANLFSPGPSQN
```

```
                        1        10         20      27
                        |--------+---------+------|
1312-0_1812-8-5_M13u    ETHITGGSQARTTRGFANLFSPGPSQN
             1312-0__   ETHITGGSQARTTRGFANLFSPGPSQN
1312-0-9_M13uni_--_1    ETHITGGSQARTTRGFANLFSPGPSQN
1312-0-4_M13uni_--_2    ETHITGGAQARTTRGFANLFSPGPSQN
1312-0-5_M13uni_--_1    ETHITGGAQARTTRGFANLFSPGPSQN
1312-0-17_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-18_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
          13120_IF-1    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-3_M13uni_--_9    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-19_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-6_M13uni_--_1    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-7_M13uni_--_4    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-13_M13uni_--_    ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-1_M13uni_--_1    ETHITGGTQARTTRGFANLFSPGTSQN
            Consensus   ETHITGGtQARTTRGFANLFSPGpSQN
```

5.3.2. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 8, 0. All amplicons correspond with an identical clone for each sample.

5.3.3 Subject D



5.3.3. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = week 16, Red = Week 12, Yellow = Week 8, Pink = Week 4, Orange = Week 0. Week 12, 8, 4, and 0 samples have the same HVR1 sequences at nucleotide level and overlap. The sample from Week 16 differs by a single amino acid substitution (see 5.3.3 Fig 2). Although the amplicon sequences predict little change, the diversity of the underlying QS milleu is not captured as illustrated by the multiple clades present in the tree.

```
                      1        10         20      27
                      |--------+---------+------|
1209-1619_M13uni_--_  HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1617_M13uni_--_  HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1618_M13uni_--_  HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1610_M13uni_--_  HTHLTGGAQARSAYQLTSLFTSGARQN
1209-161_M13uni_--_1  HTHLTGGAQARSAYQLTSLFTSGARQN
          120916_IFI  HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-1612_M13uni_--_  NTYVTGGVPSRSAHQLTSLFTSGARQN
1209-1620_M13uni_--_  RTHVTGGVPSRSAYQLTSLFTSGARQN
           Consensus  hThvTGGvqsRSAyQLTSLFTSGARQN
```

```
                        1        10         20      27
                        |--------+---------+------|
1209-1213_M13uni_--_    HTHVTGGAQARSAYQLTSLFTSGAGQN
1209-1216_M13uni_--_    HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1219_M13uni_--_    HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1220_M13uni_--_    HTHLTGGAQARSAYQLTSLFTSGARQN
1209-1215_M13uni_--_    HTHLTGGVQARSAHQLTSLFTSGARQN
1209-1210_M13uni_--_    HTYVTGGVQARSAHQLTSLFTSGARQN
1209-1217_M13uni_--_    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-125_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-121_M13uni_--_3    HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-1218_M13uni_--_    HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-1212_M13uni_--_    HTYVTGGVPSRSAYQLTSLFTSGARQN
          120912_IFI    HTYVTGGVPSRSAYQLTSLFTSGARQN
           Consensus    HTyvTGGvqaRSAyQLTSLFTSGArQN
```

```
                        1        10         20      27
                        |--------+---------+------|
1209-8-2_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-4_M13uni_--_9    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-8_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-6_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-15_M13uni_--_    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-7_M13uni_--_1    HTYVTGGVPSRSAYQLTSLFTSGARQN
           12098_IF1    HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-8-3_M13uni_--_1    RTHVTGGVPSRSAYQLTSLFTSGARQN
1209-8-11_M13uni_--_    RTHVTGGVPSRSAYQLTSLFTSGARQN
           Consensus    hTyVTGGVpSRSAYQLTSLFTSGARQN
```

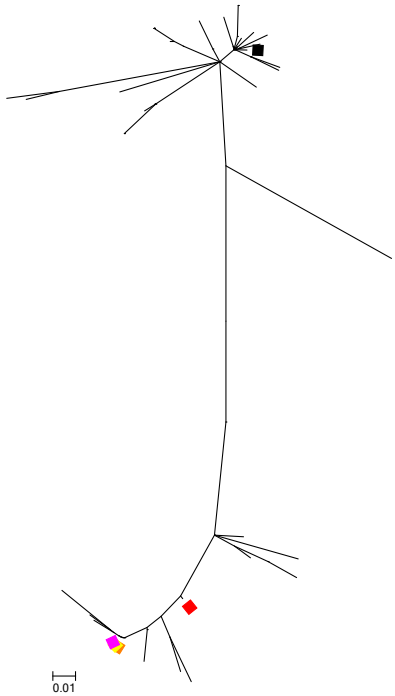
```
                        1        10         20      27
                        |--------+---------+------|
           12094_IF1    HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-4-10_M13uni_--_    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-4_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-7_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-1_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-11_M13uni_--_    HTYVTGGVPSRSAHQLASLFTSGARQN
           Consensus    HTYVTGGVpSRSAyQLtSLFTSGARQN
```

```
                        1        10         20      27
                        |--------+---------+------|
           12090_IF1    HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-0-12_M13uni_--_    HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-0-13_M13uni_--_    HTYMTGGVPSRSAYQLTSLFTSGARQN
             1209-0-6   HTYVTGGVPSRSAHQLASLFTSGARQN
1209-0-7_M13uni_--_1    HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-0-2_M13uni_--_1    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-0-11_M13uni_--_    HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-0-1_M13uni_--_1    HTYMTGGAPSRSAYQLASLFTSGARQN
           Consensus    HTYvTGGvpSRSAyQLtSLFTSGARQN
```

5.3.3. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 8, 4, 0. Each amplicon has an identical clone sequence with the exception of week 16 where the amplicon is identical to the consensus and is likely to reflect a composite sequence combining different subpopulations.

5.3.4 Subject F



5.3.4. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = Week 16, Red = Week 12, Yellow = Week 8, Pink = 4, Orange = Week 0. The Week 16 sample differs at 6 amino acid positions from the sample at Week 12 (seen in red) suggesting significant change in HVR1 during the study. The labels for samples from Weeks 8, 4, and 0 overlap as they have identical nucleotide sequence.

```
                        1        10         20      27
                        |--------+---------+------|
150566-16-150566-16-    GTHYTGGSAARGASTLAGLFTPGAQQK
         150666-16-19   GTHYTGGSAARGASTLAGLFTPGAQQK
             1505_IFI   GTHYTGGSAARGASTLAGLFTPGAQQK
           150566-16-4  GTHYTGGSAVRGASTLAGLFTPGAQQK
             Consensus  GTHYTGGSAaRGASTLAGLFTPGAQQK
```

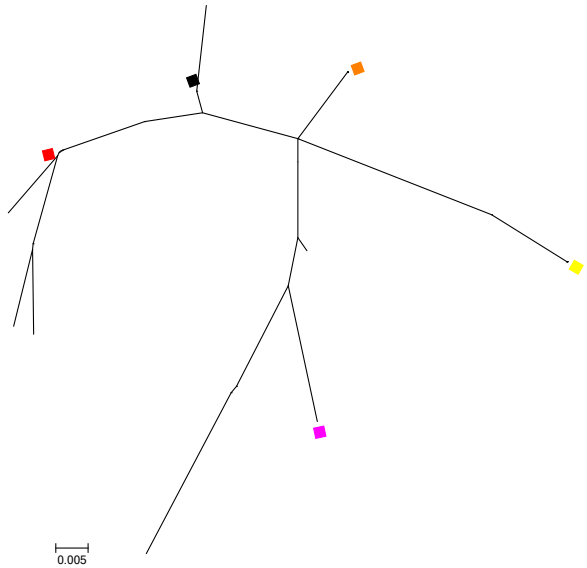
```
                        1        10         20      27
                        |--------+---------+------|
     15051215_M13uni    NTHYTGGSAARDAFRLTNLFSVGAQQK
          150512_IFI    NTHYTGGSAARDAFTLTNLFTVGAQQK
     15051212_M13uni    GTHYTGGSAAHGASTLAGLFTRGAQQK
      1505124_M13uni    GTHYTGGSAARGASTLAGLFTPGAQQK
      1505128_M13uni    GTHYTGGSAARGASTLAGLFTPGAQQK
     15051211_M13uni    GTHYTGGSAARGASTLAGLFTPGAQQK
1505123_M13uni505122    GTHYTGGSAARGASTLAGLFTPGAQQK
             Consensus  gTHYTGGSAArgAstLagLFt.GAQQK
```

```
                      1        10         20      27
                      |--------+---------+------|
150589_M13uni505813_  GTHVTGGSAARGASTLAGLFTPGAQQK
      1505815_M13uni  GTHVTGGSAARGASTLAGLFTPGAQQK
       150585_M13uni  GTHVTGGSAAHGASTLAGLFTRGAQQK
       150582_M13uni  GTHVTGGSAARAASGLAGLFSPGAQQK
       150587_M13uni  NTHVTGGAAARDAFGLANLFSVGAQQK
      1505810_M13uni  NTHVTGGSAARDAFRLTNLFSVGAQQK
            15058_IFI NTHVTGGSAARDAFRLTNLFSVGAQQK
            Consensus gTHVTGGsAAr.As.LagLFs.GAQQK


                      1        10         20      27
                      |--------+---------+------|
           15054_IF1  NTHVTGGSAARDAFRLTNLFSVGAQQK
1505-4-9_M13uni_--_2  NTHVTGGSAARDAFRLTNLFSVGAQQK
1505-4-1_M13uni_--_5  NTHITGGSAARDAIRLTNLFSVGAQQK
1505-4-12_M13uni_--_  GTHVTGGSAARGASTLAGLFTPGAQQK
1505-4-18_M13uni_--_  GTHVTGGSAARAASTFAGLFTPGAQQK
1505-4-19_M13uni_--_  GTHVTGGSAAHGASTLAGLFTRGAQQK
1505-4-20_M13uni_--_  DTHVTGGSVARGASAFAGLFSQGAQQK
1505-4-3_M13uni_--_3  DTHVTGGSVARAASSFAGLFSQGAQQK
            Consensus .TH!TGGSaAr.As.lagLFs.GAQQK


                      1        10         20      27
                      |--------+---------+------|
                 x0-17 GTHVTGGSAARDAFRLTNLFSVGAQQK
           15050_IF1  NTHVTGGSAARDAFRLTNLFSVGAQQK
                 x0-16 NTHVTGGSAARDAFRLTNLFSVGAQQK
                  X0-8 NTHVTGGSAARDAFRLTNLFSVGAQQK
                  X0-7 GTYVTGGSAARAASTFAGLFTPGAQQK
                  X0-1 GTHVTGGSAARGASTLAGLFTPGAQQK
                  X0-2 GTHVTGGSAARGASTLAGLFTPGAQQK
                  X0-6 GTHVTGGSAAHGASTLAGLFTRGAQQK
            Consensus gThVTGGSAAr.AstlagLFt.GAQQK
```

5.3.4. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 8, 4, 0. Each amplicon sequence is identical to one cloned sequence from the equivalent sample.

5.3.5. Subject K



5.3.5. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = Week 16, Red = Week 12, Yellow = Week 8, Pink = Week 4, Orange = Week 0.

```
                        1       10        20     27
                        |--------+---------+------|
          010516_IF-1   GTRISGGSAAHTINRFTGLFSSGPSQK
   E1612_M13uni-21_--   GTHISGGSAAHTINRFTGLFSSGPSQK
E14-3_M13uni-21_--_7    GTHISGGSAARTINRFTGLFSSGPSQK
E16-9_M13uni-21_--_9    ETYISGGSAAHTVSRFTGLFSSGPSQK
            Consensus   gT.ISGGSAAhT!nRFTGLFSSGPSQK
```

```
                        1       10        20     27
                        |--------+---------+------|
E-12-1_M13uni_--_20.    GTHISGGSAARTINRFTGLFSSGPSQK
          010512_IF-1   GTHISGGSAARTINRFTGLFSSGPSQK
E-12-18_M13uni_--_10    GTHISGGSAARTINRFTGLFSSDPSQK
            Consensus   GTHISGGSAARTINRFTGLFSSgPSQK
```

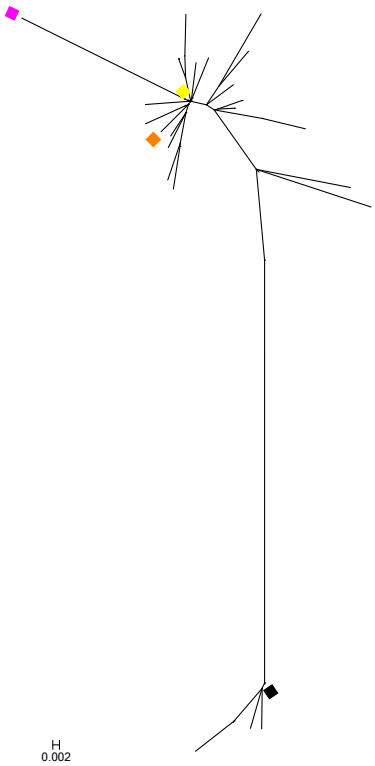
```
                        1       10        20     27
                        |--------+---------+------|
0105-8-9_M13uni_--_2    GTHISGGSAARTINRFTGLFSSGPSQK
0105-8-16_M13uni_--_    GTHISGGSAARTINRFTGLFSSGPSQK
0105-8-18_M13uni_--_    GTHISGGSAARTINRFTGLFSSGPSQK
           01058_IF-1   GTRISGGSAAYATSRFTGLFSSGPSQK
            Consensus   GThISGGSAArtinRFTGLFSSGPSQK
```

```
                        1        10         20      27
                        |--------+---------+------|
E4-2_M13uni-21_--_42    GTRISGGSAAYATSRFTGLFSSGPSQK
E4-16_M13uni-21_--_2    GTRISGGSAAHTVSRFTGLFSSGPSQK
          01054_IF-1    ETYISGGSAAHATSRFTGLFSSGPSQK
           Consensus    gTrISGGSAAhatSRFTGLFSSGPSQK


                        1        10         20      27
                        |--------+---------+------|
           0105-0-10    GTRISGGSAAYATSRFTGLFSSGPSQK
          01050_IF-1    GTRISGGSAAHTVSRFTGLFSSGPSQK
           Consensus    GTRISGGSAAhatSRFTGLFSSGPSQK
```

5.3.5. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 8, 4, 0.

5.3.6. Subject L



5.3.6. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 1000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = Week 16, Yellow = Week 8, Pink = Week 4, Orange = Week 0. The sequence at Week 16 differs at 9 amino acid positions from the amplicon sequence seen at Week 8 and maps remotely on the phylogenetic tree. However, the cloning sequences include identical sequences to those seen later in the study suggesting the presence of multiple subpopulations (see 5.3.6. Fig 2)

```
                          1       10        20      27
                          |--------+---------+------|
          190716_IF-1     VTHVTGGSVAREAFGLTSLFSRGAQQK
     1907-16-5_M13uni     VTHVTGGSVAREAFGLTSLFSRGAQQK
1907-16-19_M13uni907      VTHVTGGSVAREAFGLTSLFSRGAQQK
1907-16-10_M13uni907      VTHVTGGSVAREAFGLTSLFSRGAQQK
1907-16-2_M13uni907-      VTHVTGGSVAREAFGLTSLFSRGAQQK
1907-16-3_M13uni907-      VTHVTGGSVAREAFGLTSLFSRGAQQK
1907-16-11_M13uni907      KTYVTGGSVAHAARGFTNLFSRGAQQN
     1907-16-7_M13uni     KTYVTGGSVAHAARGFTNLFSRGAQQK
            Consensus     vThVTGGSVAreAfGlTsLFSRGAQQk
```

```
                        1        10         20      27
                        |--------+---------+------|
1907-8-8_M13uni_--_5    KTYVTGGSVAHAARGFTNLFSRGAQQN
1907-8-5_M13uni_--_7    KTYVTGGSVAHAARGFTNLFSRGAQQN
           19078_IF-1   KTYVTGGSVAHAARGFTNLFSRGAQQN
1907-8-4_M13uni_--_8    KTYVTGGSVAHVARGFTNLFSRGAQQN
            Consensus   KTYVTGGSVAHaARGFTNLFSRGAQQN
```

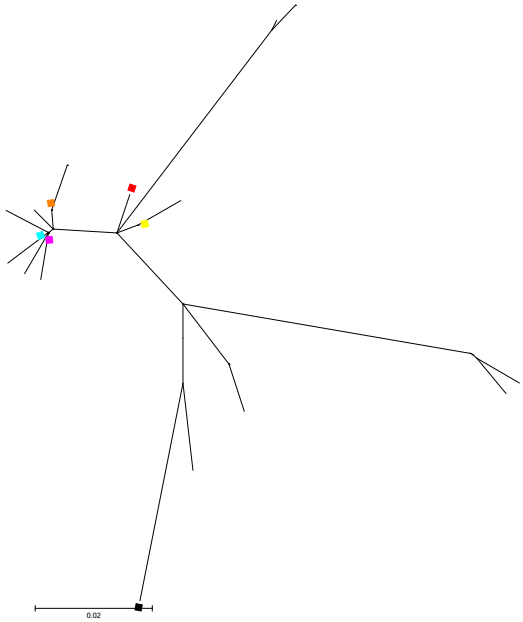
```
                        1        10         20      27
                        |--------+---------+------|
1907417_M13uni907416    KTYVTGGSVAHAARGFTNLFSRGAQQN
      1907420_M13uni    KTYVTGGSVAHAARGFTNLFSRGAQQN
      1907415_M13uni    KTYVTGGSVAHAARGFTNLFSRGAQQN
      1907419_M13uni    KTYVTGGSVAHVARGFTNLFSRGAQQN
          19074_IF-1    KTYVTGGSVAHAARAFTNLFSSGAHQN
            Consensus   KTYVTGGSVAHaARgFTNLFSrGAqQN
```

```
                        1        10         20      27
                        |--------+---------+------|
1907-0-1907-0-2907-0    NTYVTGGSVAHAARGFTNLFSRGAQQN
1907-0-7907-0-13907-    NTYVTGGSVAHAARGFTNLFSRGAQQN
           19070_IF-1   NTYVTGGSVAHAARGFTNLFSRGAQQN
             1907-0-12  NTYVTGGSVAHVARGFTNLFSRGAQQN
             1907-0-19  TTYVTGGSVAHAARGFTNLFSRGAQQN
              1907-0-6  KTYVTGGSVAHAARGFTNLFSRGAQQN
             1907-0-16  KTYVTGGSVAHAARGFTNLFSRGAQQN
              1907-0-8  NTYVTGGSVAHAARGLTNLFSRGAQQN
            Consensus   nTYVTGGSVAHaARGfTNLFSRGAQQN
```

5.3.6. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 6, 4, 0.

5.3.7. Subject M



5.3.7. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = Week 16, Red = Week 12, Yellow = Week 8, Pink = Week 4, Turquoise = Week 2, Orange = Week 0. The Week 16 amplicon sequence differs from Week 12 by 6 amino acid substitutions suggesting temporal change.

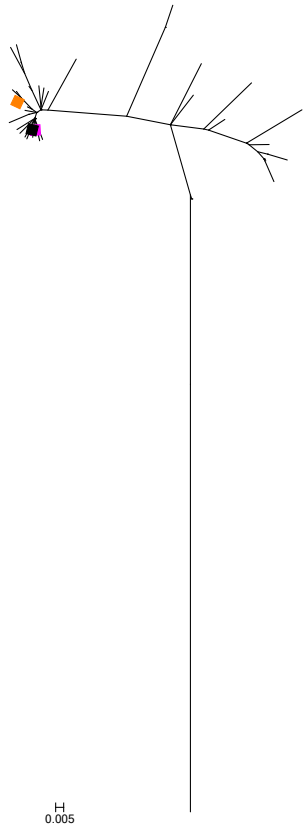
```
                      1        10         20      27
                      |--------+---------+------|
150981612_M13uni5091  ATYITGGASAHTTYGLASLFTAGPRQK
    150916841_M13uni  ATYITGGASAHTTYGIASLFTAGPRQK
    150916810_M13uni  ATYITGGASAHATYGLASLFTAGPRQK
15091687_M13uni50916  TTRITGGASAHNTYGLASLFSVGAQQK
           150916_IFI TTRITGGASAHNTYGLASLFSVGAQQK
            Consensus aTyITGGASAH.TYGlASLFtaGprQK

                      1        10         20      27
                      |--------+---------+------|
      1509121_M13uni  TTRISGGASAHNTYGLASLFSVGAQQK
     15091214_M13uni  TTRITGGASAHNTYGLASLFSVGAQQK
      1509122_M13uni  TTYISGGASAHTTYGLASLFTAGAQQK
      1509124_M13uni  ATYITGGASAHTTYGLASLFTAGPRQK
1509127_M13uni509121  ATYITGGASAHTTYGLASLFTAGPRQK
           150912_IF1 ATYITGGASAHTTYGLASLFTAGPQQK
     15091215_M13uni  ATYITGGASAHTTYGLTSLFTAGPRQK
            Consensus tTyItGGASAHtTYGLaSLFtaGaqQK
```

```
                        1        10         20      27
                        |--------+---------+------|
150981_M13uni50982_M    TTRITGGASAHNTYGLASLFSVGAQQK
       150984_M13uni    TTRITGGASAHNTYGLASLFSVGAQQK
150983_M13uni50987_M    ATYITGGASAHTTYGLASLFTAGPRQK
           15098_IF1    ATYITGGASAHTTYGLASLFTAGPRQK
      1509817_M13uni    TTYITGGASAHTTYGLASLFTAGPRQK
           Consensus    tTyITGGASAHtTYGLASLFtaGprQK


                        1        10         20      27
                        |--------+---------+------|
           15094_IF1    ATYITGGASAHTTYGLASLFTAGPRQK
150943_M13uni50944_M    ATYITGGASAHTTYGLASLFTAGPRQK
150942_M13uni50947_M    ATYITGGASAHTAYGLASLFTAGPRQK
           Consensus    ATYITGGASAHTtYGLASLFTAGPRQK


                        1        10         20      27
                        |--------+---------+------|
G2-1_M13uni-21_--_23    ATYITGGASAHTTYGLASLFTAGPRQK
G2-8_M13uni-21_--_50    ATYITGGASAHTTYGLASLFTAGPRQK
           15092_IF1    ATYITGGASAHTTYGLASLFTAGPRQK
G-2-192-16_M13uni-21    ATYITGGASAHTAYGLASLFTAGPRQK
G2-10_M13uni-21_--_1    TTYITGGASAHTTFGLASLFTKGAQQK
           Consensus    aTYITGGASAHTt%GLASLFTaGprQK


                        1        10         20      27
                        |--------+---------+------|
15090450905509065090    ATYITGGASAHTTYGLASLFTAGPRQK
          15090_IF-1    ATYITGGASAHTTYGLASLFTAGPRQK
           Consensus    ATYITGGASAHTTYGLASLFTAGPRQK
```

5.3.7. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 6, 4, 2, 0.

5.3.8. Subject N



5.3.8. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = week 16, Pink = Week 2, Orange = Week 0. The labels overlap as the amplicon sequences are almost identical with the sequences from Week 16 and 2 differing from Week 0 by a single amino acid substitution.

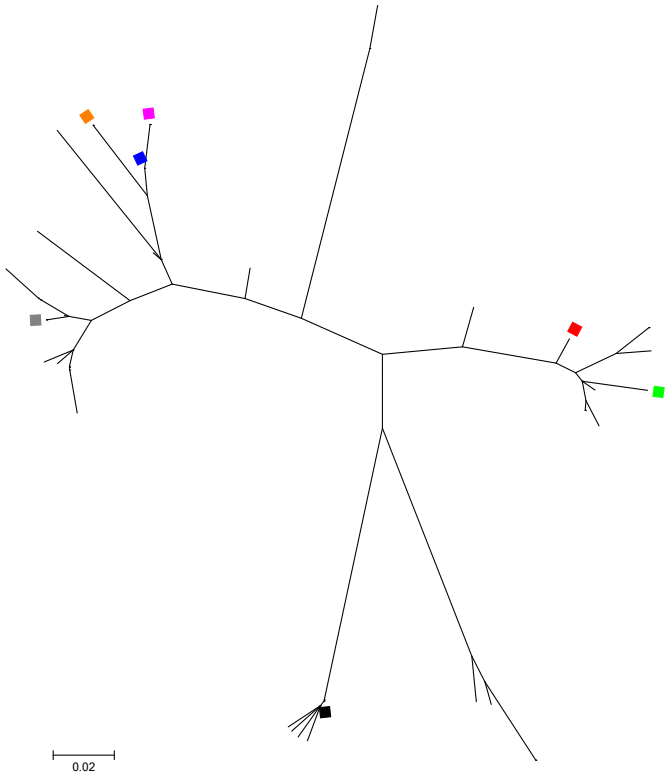
```
                        1       10        20      27
                        |--------+---------+------|
          230816_IF-1   ETHVTGGTAAHGALGITSLLSRGPKQN
62_M13uni7_M13uni5_M    ETHVTGGTAAHGALGITSLLSRGPKQN
            68_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            80_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            74_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            69_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            61_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            64_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
79_M13uni8_M13uni6_M    ETHVTGGTAAHGALGITSLLSRGPKQN
            71_M13uni   ETHVTGGTAAHGALGITSLLSPGPKQN
            77_M13uni   ETHVTGGTTAHGAFGITSLLSRGPKQN
            70_M13uni   ETHVTGGTAAHGVFGITSLLSSGPKQN
   63_M13uni6_M13uni    ETHVTGGTAAHGAFTLTTLLNRGAKQN
            73_M13uni   ETHVTGGTAAHGAFTLTTLLNRGAKQN
            75_M13uni   DTRVTGGTVAHGALTLTSLLSRGAKQN
            Consensus   #ThVTGGTaAHGalgiTsLLsrGpKQN
```

```
                        1        10         20      27
                        |--------+---------+------|
2308-2-1_M13uni_--_1    ETHVTGGTAVHGALGITSLLSRGPKQN
2308-2-10_M13uni_--_    ETHVTGGTAAHGALGITSLLSRGPKQN
2308-2-21_M13uni_--_    ETHVTGGTAAHGALGITSLLSRGPKQN
           23082_IF-1   ETHVTGGTAAHGALGITSLLSRGPKQN
2308-2-16_M13uni_--_    ETHVTGGTAAHGVFGITSLLSRGPKQN
2308-2-6_M13uni_--_8    ETHVTGGTAAHGALGISRLLSLGPKQN
2308-2-19_M13uni_--_    DTRVTGGTVAHGALTLTSLLSRGAKQN
2308-2-20_M13uni_--_    ETHVTGGTAAHGAFTLTTLLSRGAKQN
2308-2-23_M13uni_--_    ETHVTGGTAAHGAFTLTTLLNRGAKQN
            Consensus   #ThVTGGTaAHGalgitsLLsrGpKQN


                        1        10         20      27
                        |--------+---------+------|
2308-0308-0-91_M13un    ETHVTGGTAAHGALGITSLLSRGPKQN
    2308-0-88_M13uni    ETHVTGGTAAHGALGITSLLSRGPKQN
    2308-0-85_M13uni    ETHVTGGTAAHGALGITSLLSRGPKQN
    2308-0-86_M13uni    ETHVTGGTVAHGALGITSLLSRGPKQN
 2308-0-62_M13uni(2)    ETHVTGGTTAHGALGITSLLSRGPKQN
           23080_IF-1   ETHVTGGTAAHGAFGITSLLSRGPKQN
    2308-0-90_M13uni    ETHVTGGTAAHGALGITSLLSRGPNQN
    2308-0-83_M13uni    ETHVTGGTAAHGVFGITSLLSPGPKQN
 2308-0-61_M13uni(2)    ETHVTGGTAAHGAFTLTTLLNRGAKQN
 2308-0-63_M13uni(2)    ETHVTGGTAAHGAFTLTTLLNRGAKQN
2308-0-94_M13uni308-    ETHVTGGTAAHGAFTLTTLLNRGAKQN
    2308-0-84_M13uni    ETHVTGGTAAHGAFTLTTLLSRGAKQN
    2308-0-87_M13uni    ETHVTGGTAAHGAFTLTTLLSRGAKQN
    2308-0-96_M13uni    ETHVTGGTAAHGAFTLTTLLSRGPKQN
    2308-0-81_M13uni    ETHVTGGTAAHGALTLTSLLSRGAKQN
            Consensus   ETHVTGGTaAHGafgiTsLLsrGpKQN
```

5.3.8. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 2, 0.

5.3.9. Subject Q



5.3.9. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = Week 16, Grey = Week 14, Red = Week 12, Green = Week 10, Blue = Week 6, Pink = Week 4, Orange = Week 0. Each amplicon sequence maps to a distinct clade within the overall tree suggesting a time order phylogeny.

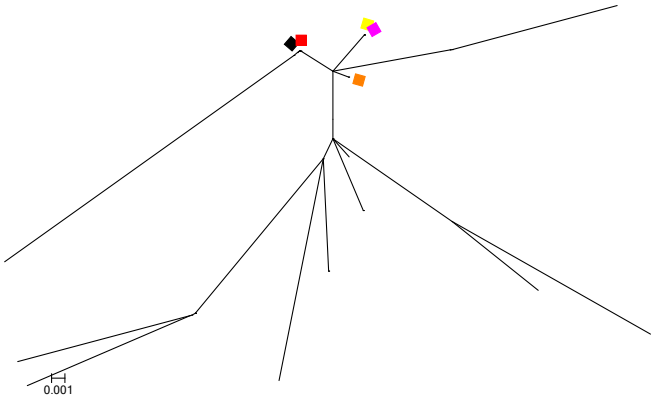
```
                      1        10         20      27
                      |--------+---------+------|
101086-16-1001086-16  ETTISGGSAARSVYTFTGFFSPGAKQK
101086-16-101086-16-  ETTISGGSAARSVYTFTGFFSPGAKQK
             1010_IFI ETTISGGSAARSVYTFTGFFSPGAKQK
M-16-13_M13uni_--_19  ETTISGGSAARSVYTFTGFFSPGAKQK
          101086-16-4 ETTISGGSAARSVYTFTGFFSPGAKQK
M-16-8_M13uni_--_9..  ETTISGGSAARSVYTFTGFFSPGAKQK
          101086-16-9 ETTITGGTAARSVSTFTKLFAPGASQK
101086-16-1101086-16  ATIITGGTAAHSVRTIADWFAPGASQK
M-16-1_M13uni_--_8..  ATIITGGTAAHSVRTIADWFAPGASQK
M-16-9_M13uni_--_647  ETTISGGSAAHSVRTIADWFAPGASQK
M-16-17_M13uni_--_7.  TTTITAGTAARGARSFANLFSPGASQN
           Consensus  eTtItgGtAArsvrtfa..FaPGAsQk
```

```
                        1        10         20      27
                        |--------+---------+------|
1010-1418_M13uni_--_    ETTITGGTAARSVSTFTRLFAPGASQK
1010-1416_M13uni_--_    ETTITGGTAAHSVRTIAEWFAPGASQK
 1010-1410_M13uni_--    TTTITAGTAARGARSFVNLFSPGASQN
  1010-141_M13uni_--    ETTISGGAAARGVHSFTSLFTPGASQK
  1010-1411_M13uni_-    ETTISGGAAARGVHSFTSLFTPGARQN
          101014_IF-1   ETTISGGAAARGAYSLTSLFTSGARQN
            Consensus   eTTItgGtAArgv.sft.lF.pGAsQn


                        1        10         20      27
                        |--------+---------+------|
M-10-22_M13uni_--_19    ETTISGGAVARGVHSFTSLFSTGARQN
M-10-10_M13uni_--_21    ETTITGGTVARGVSTITRLFAPGASQK
          101010_IF-1   ETTITGGTAARDVSTFTRLFAPGASQK
            Consensus   ETTItGGtvARgVstfTrLFapGAsQk


                        1        10         20      27
                        |--------+---------+------|
M-6-10_M13uni_--_8..    ETTISGGAAARSAYSLTSLFTPGARQN
           10106_IF-1   ETTISGGAVARGVHSFTSLFTSGARQN
            Consensus   ETTISGGAaARgahSlTSLFTpGARQN


                        1        10         20      27
                        |--------+---------+------|
10108677_M13uni01086    ETTISGGAAARGAYSLTNLFTPGARQN
     10108678_M13uni    ETTISGGAAARGAYSLTNLFTPGARQN
     10108666_M13uni    ETTITGGAAARGAYSLTNLFTPGARQN
           10104_IF1    ETTISGGAVARGVHSFTSLFKSGARQN
            Consensus   ETTIsGGAaARGaySlTnLFtpGARQN


                        1        10         20      27
                        |--------+---------+------|
10108693_M13uni01086    ETTISGGAAARSAYSLTSLFTSGARQN
           10100_IF1    ETTISGGAVARGVHSFTSLFSTGARQN
10108682_M13uni01086    ETTISGGAVARGVHSFTSLFKSGARQN
            Consensus   ETTISGGAvARgvhSfTSLF.sGARQN
```

5.3.9. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 14, 12, 10, 6, 4, 0.

5.3.10. Subject T



5.3.10. Fig 1.

Maximum likelihood composite phylogenetic tree with boot strap of 10,000 comprising all unique nucleotide HVR1 sequences from prospective cloning project (unlabelled) with the amplicon sequences included and labelled. Black = Week 16, Red = Week 12, Yellow = Week 8, Pink = Week 4, Orange = Week 0. All amplicon sequences are identical at amino acid level and differ by a maximum of one nucleotide substitution from each other.

```
                        1       10        20      27
                        |--------+---------+------|
          300716_IFI    ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007163_M13uni007162    ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007161_M13uni007166    ETRVTVGQAAYTTNAVTRFFSPGAKQN
           Consensus    ETRVTVGQAAYTTNAVTRFFSPGAKQN


                        1       10        20      27
                        |--------+---------+------|
          300712_IF1    ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007-1217_M13uni_--_    ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007-128_M13uni_--_1    ETRVTVGQAAYTTNAVTRFFSPGAKQN
           Consensus    ETRVTVGQAAYTTNAVTRFFSPGAKQN


                        1       10        20      27
                        |--------+---------+------|
           30078_IF1    ETRVTVGQAAYTTNAVTRFFSPGAKQN
300787_M13uni007816_    ETRVTVGQAAYTTNAVTRFFSPGAKQN
300785_M13uni00789_M    ETRVTVGQAAYTTNAVTRFFSPGAKQN
           Consensus    ETRVTVGQAAYTTNAVTRFFSPGAKQN
```

```
                        1       10        20     27
                        |--------+---------+------|
            30074_IF1   ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007-4-3_M13uni_--_1    ETRVTVGQAAYTTNAVTRFFSPGAKQN
V-4-10_M13uni_--_17.    ETRVTVGQAAYTTNAVTRFFSPGAKQN
V-4-1_M13uni_--_11..    ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007-4-1_M13uni_--_2    ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007-4-12_M13uni_--_    ETRVTVGQAAYTTNAVTRFFSPGAKQN
V-4-16_M13uni_--_12.    ETRVTVGQAAYTTNVVTRFFSPGAKQN
            Consensus   ETRVTVGQAAYTTNaVTRFFSPGAKQN


                        1       10        20     27
                        |--------+---------+------|
            30070_IF1   ETRVTVGQAAYTTNAVTRFFSPGAKQN
3007010_M13uni007017    ETRVTVGQAAYTTNAVTRFFSPGAKQN
              3007019   ETRVTVGQAAYTTNAVTRFFSPGAKQN
       300703_M13uni0   ETRVTVGQAAYTTNAVTRFFSPGAKQN
            Consensus   ETRVTVGQAAYTTNAVTRFFSPGAKQN
```

5.3.10. Fig 2.

Amino acid multiple sequence alignments of the unique nucleotide HVR1 sequences for each sample from the cloning project combined with the corresponding amplicon sequence. The alignments were constructed using multalin and the figure includes samples from weeks 16, 12, 8, 4, 0.

5.3.11. Results Summary

Subject	Sample (weeks)	Amino Acid differences between Amplicon and Master sequence	Nucleotide differences between Amplicon and Nearest Clone
B	16	2	5
	12	0	0
	6	0	0
	4	0	0
	0	0	0
C	16	0	1
	12	0	1
	8	0	1
	0	0	1
D	16	3	3
	12	0	0
	8	0	0
	4	0	2
	0	0	0
F	16	0	1
	12	2	2
	8	0	0
	4	8	0
	0	8	0
K	16	1	1
	12	1	1
	8	5	8
	4	2	4
	0	3	4

Subject	Sample (weeks)	Amino Acid differences between Amplicon and Master sequence	Nucleotide differences between Amplicon and Nearest Clone
L	16	0	0
	8	0	0
	4	3	3
	0	0	0
M	16	0	0
	12	1	1
	8	0	0
	4	0	0
	2	0	0
	0	0	0
N	16	0	0
	2	0	0
	0	1	0
Q	16	0	2
	14	4	4
	10	3	3
	6	6	6
	4	7	7
	0	2	4
T	16	0	0
	12	0	0
	8	0	0
	4	0	0
	0	0	2

5.3.11. Table 1

The number of amino acid and nucleotide differences between the amplicon sequence and the master sequence (most prevalent sequence generated by the cloning project for the equivalent sample).

	Genotype	Cirrhosis	Time ordered phylogeny	Identified by Amplicon	Single Dominant Subpopulation	Identified by Amplicon	Replacement of dominant subpopulation	Identified by Amplicon
B	1b	Y		Y	+	Y		Y
C	1b	Y		Y	+	Y		Y
D	1b	Y		Y		Y		Y
F	1b	N	+	N		Y	+	N
K	3a	N	+	Y		Y	+	Y
L	3a	N		Y		Y	+	Y
M	3a	N		Y		Y	+	Y
N	3a	N		Y	+	Y		Y
Q	3a	N	+	Y		Y	+	Y
T	3a	Y		Y	+	Y		Y

5.3.11. Table 2

Predictive power of amplicon sequencing for changes seen in the clonal analysis.

Amplicon sequencing corresponded with the cloning project in correctly identifying all subjects who had a single dominant HVR1 QS subpopulation. In predicting the presence or absence of time order

phylogeny there was concordance between the amplicon and cloning data in nine out of ten subjects. Finally, the amplicon sequences confirmed the replacement of the initial dominant HVR1 QS in nine out of ten subjects. Amplicon and cloning findings differed for subject F where the cloning data illustrated a complex interaction with two prominent disparate subpopulations. Cloning data suggested changes in the prevalence of these subpopulations with time but the amplicon sequences suggest both a change in the dominant subpopulation and a time order phylogeny.

Nucleotide Substitution	No.
G-A	9
A-G	8
T-G	1
G-T	2
T-C	18
C-T	13
C-G	4
G-C	3
A-C	3
C-A	3
T-A	2
A-T	1

5.3.11. Table 3.

Collated data for all of the amplicon project indicating the patterns of nucleotide differences between the amplicon and the most prevalent cloning sequence. 72% of differences are the result of a purine to purine substitution or a pyrimidine to pyrimidine substitution.

Subject	Number of Subpopulations	Master Sequence within number of Amino acid Substitutions of Amplicon							
		1	2	3	4	5	6	7	8
N	1	100	100	100	100	100	100	100	100
T	1	100	100	100	100	100	100	100	100
D	1	80	80	100	100	100	100	100	100
L	2	75	75	100	100	100	100	100	100
M	3	100	100	100	100	100	100	100	100
B	3	80	100	100	100	100	100	100	100
K	3	40	60	80	80	100	100	100	100
F	5	40	60	60	60	60	60	60	100
Q	7	17	33	50	67	67	73	100	100

5.3.11. Table 4.

Concordance of master sequence as predicted by amplicon and nucleotide sequences stratified according to how many subpopulations were identified in the cloning analysis. The number of subpopulations was associated with an inverse likelihood of the cloning data and amplicon data identifying the same master sequence.

5.4. Discussion

We investigated how representative the sequencing of the HVR1 PCR amplicon was of the underlying HVR1 QS as demonstrated using cloned data. A mean of 17 (range 12-24) sequences generated for each sample using cloning techniques were compared with the amplicon data. Although many studies have investigated change in HVR1 quasispecies, there is no published literature either describing changes in sequenced amplicon products or comparing amplicon sequencing with cloned quasispecies profiles(308, 309, 314, 338, 343, 346).

Interestingly, the amplicon sequence correctly predicted the most dominant cloned sequence as identified in 63% of samples. This highlights an ability to identify QS change. Conversely, the amplicon failed to identify the dominant clonal sequence in 37% of cases. The amplicon sequence was within 2 or 3 amino acid substitutions of the master sequence in 80% and 88% of samples respectively.

When we compared the amplicon sequences with cloning data, the amplicon correctly identified 80% of the subjects where there was a change in the dominant HVR1 sequence during the study.

In cases where the cloned sequences identified a single dominant subpopulation, the amplicon sequence correctly identified a sequence that would be contained within that subpopulation in 100% of cases. However, the lack of concordance between cloning data and amplicon sequencing has important implications for the new direct anti viral medications as currently all screening for drug resistance utilises Sanger sequencing of the virus. Although HVR1 quasispecies represents an extreme in the depth of viral diversity, our data raises questions with regard to the reliability of Sanger sequencing for the purposes of screening for viral resistance.

We investigated for possible explanations for this inability to identify the most prevalent clone in our cloning data. We identified an association between multiple subpopulations in the cloning samples and a greater risk of disparity between the amplicon and cloning sequences ($p<0.05$ *Mann Whitney U*). In a number of cases where the amplicon did not correspond with any of the cloned sequences, this was because the amplicon represented a composite of the two co dominant subpopulations (Subject D – week 16 -7.3.3. Fig. 2 and Subject E – week 12 – 7.3.4. Fig. 2).

When we compared the amplicon nucleotide sequence generated with the closest cloned sequence we identified the identical sequence in 50% of samples. We examined the sequences in order to identify the most common substitutions in the closest cloning sequence to the amplicon and found that it was a transition in 31% of cases and a transversion in 69%.

The most common substitutions were substitutions of a G-A/A-G or C-T/T-C. Substitution of a purine for a purine or pyrimidine for pyrimidine constituted 72% of all differences seen between the amplicon sequence and the next nearest clone sequence.

We generated our amplicon sequences using *Taq* polymerase, a DNA polymerase which lacks a proofreading function and has an estimated error rate of 2×10^{-4} to $>1 \times 10^{-5}$ errors/site/cycle. The cloning sequences were generated using polymerase chain reactions catalysed by *Pwo* polymerase, a DNA polymerase with a proofreading mechanism that has an estimated error rate ten times less than *Taq* polymerase(404). The use of *Taq* polymerase would, we calculate, result in 0.03-0.07 errors per sequence in the 81 base pair sequence corresponding with the HVR1 after two PCR cycles(405). Importantly, the likelihood that a missubstitution would occur early enough in the PCR cycle such that it could be represented as dominant in the final PCR product is extremely remote given the high numbers of circulating virus sequences in the pre reverse transcriptase sample. Hence we are satisfied that the discrepancy noted between the amplicon sequences and the clones generated are not as a result of the DNA polymerase used.

5.5. Conclusion

Amplicon sequencing can correctly identify the most dominant sequence in most cases and can also suggest HVR1 QS undergoing significant change. However, in subjects where there is high HVR1 diversity as seen in cases with multiple subpopulations, the ability of amplicon sequencing to identify the dominant QS is more limited. Although amplicon sequencing can be used as a blunt tool to identify HVR1 change, it does not provide sufficient surrogate information with regard to the underlying QS to obviate the use of cloning. Finally, discrepancy between cloning and amplicon sequencing may be a harbinger of future challenges in screening for resistance to new direct anti virals.

Chapter 6

Analysis of long term HVR1 sequence evolution

6.1 Background

Short interval change in HCV HVR1 complexity and diversity is unpredictable.

Our analysis also identified the contrasting tempo of HVR1 change from stasis to time order phylogenetic change over intervals far shorter than had previously been described.

In order to further investigate whether the QS stasis/change patterns we described are sustained over more prolonged periods of time we cloned and sequenced a retrospective sample from the HCV library curated by the Molecular Virology Diagnostics and Research Laboratory where a suitable sample was available (n=18 of the 23 subjects included in the prospective study).

This strategy would allow us to investigate the tempo of HVR1 change over a more prolonged period of time in order to clarify:

1. Among those subjects where we identified time order phylogeny whether this pattern has persisted in the form of novel subpopulations in the retrospective sample.
2. Whether HVR1 stasis as described is sustained over prolonged periods.
3. Although, the clonal depth used has been suggested to be sufficient for analysis of quasispecies change, it is possible that unavoidable random selection bias may have distorted the patterns we identified and the inclusion of a retrospective sample may facilitate the identification of subpopulations identified late in the prospective cloning study(323).

Finally, we demonstrated highly variable rates of HVR1 QS change in the substitution per site per year but also rates that were 10 fold greater than previous studies of HCV had suggested(406, 407). The effect of the subpopulations within a quasispecies milieu on estimated substitution rates is unclear. The inclusion of a temporally remote sample would permit the confirmation of HCV HVR1 substitution rates.

6.2 Methods

Described in Methods Chapter section 2.5.

6.3 Results

The results for four of the five subjects presented in the prospective cloning study are presented individually (no suitable sample was available for subject T). These subjects provide illustration of the patterns of change we identified.

The remaining subjects are presented separately in Appendix B.

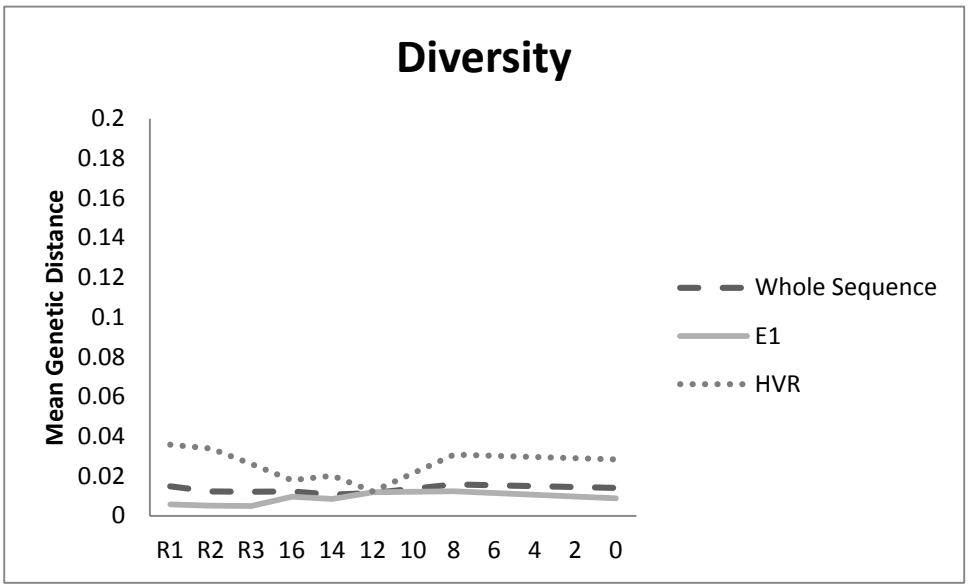
The results are presented separately in the following order:

1. Diversity, complexity and divergence
2. Phylogenetic change
3. Subpopulation analysis

A summary of the overall data for all 18 subjects is presented in section 6.3.5.

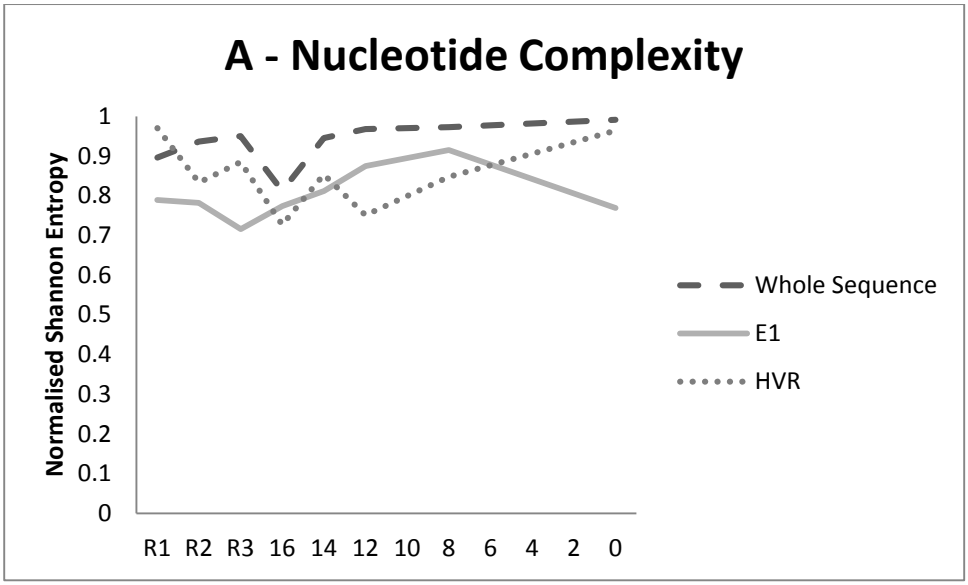
6.3.1 Subject C

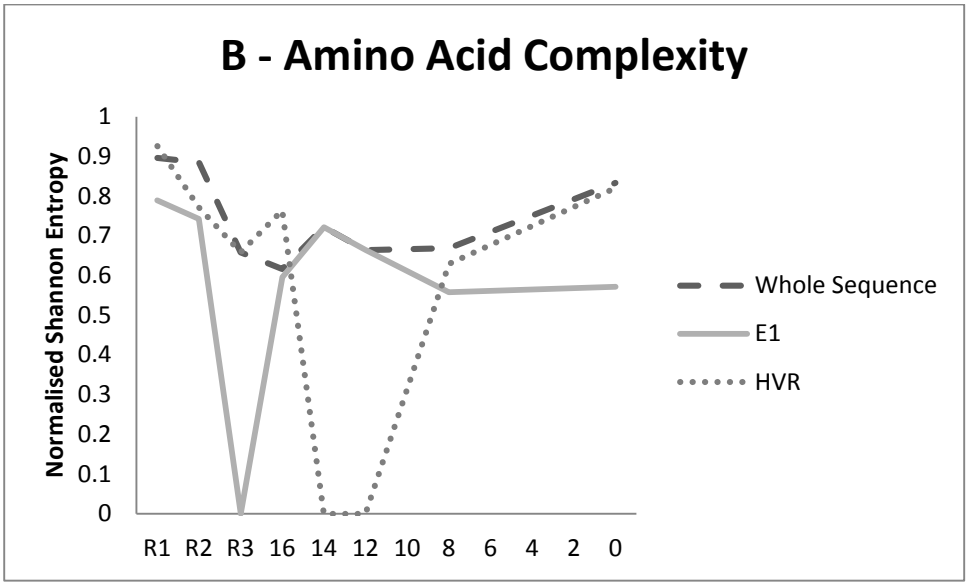
6.3.1 Diversity, Complexity, and Divergence



6.3.1 Fig 1. HVR1 QS diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

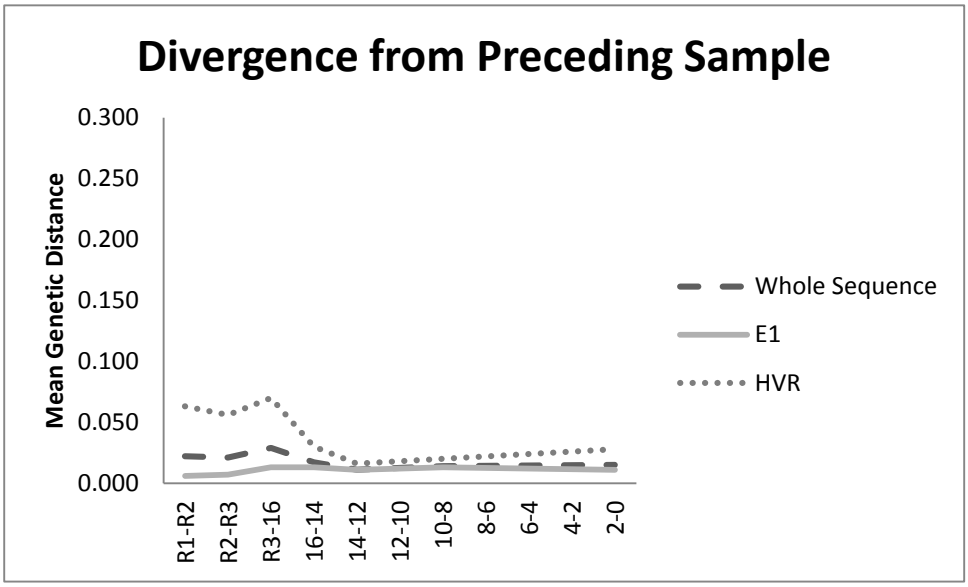
HVR1 diversity is greater than E1 diversity, however overall diversity is low for the entirety of the study. This may imply little change of a homogenous QS.





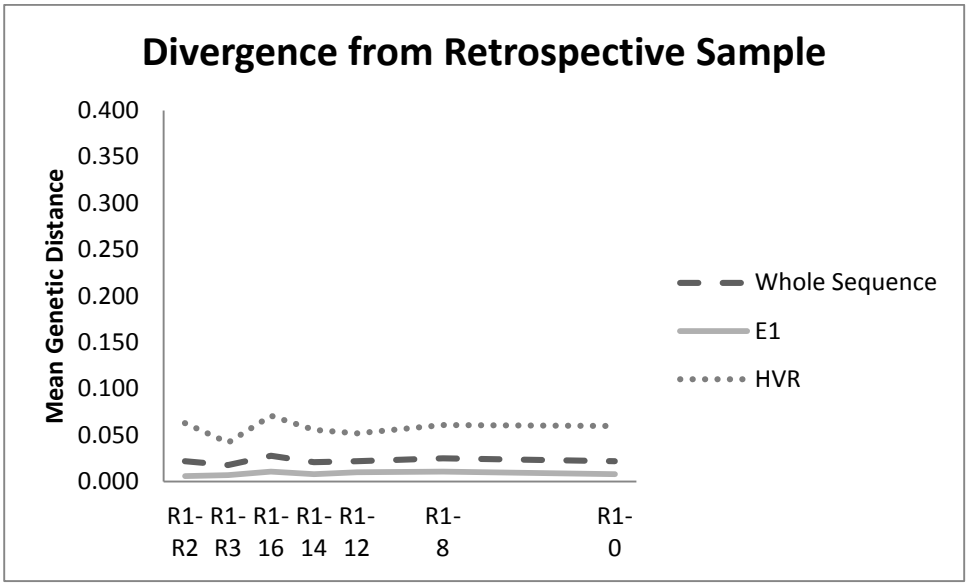
6.3.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy. High nucleotide complexity reflects in the context of the low diversity seen in 6.3.1 Fig 1 suggests a QS that is confined to a local fitness optimum. The variable amino acid complexity indicates the variable appearance of closely related sequences within the clones.

HVR1 demonstrates greater nucleotide complexity relative to E1 in most samples though this is not the case for amino acid complexity which suggests a dominant master sequence under purifying selection.



6.3.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

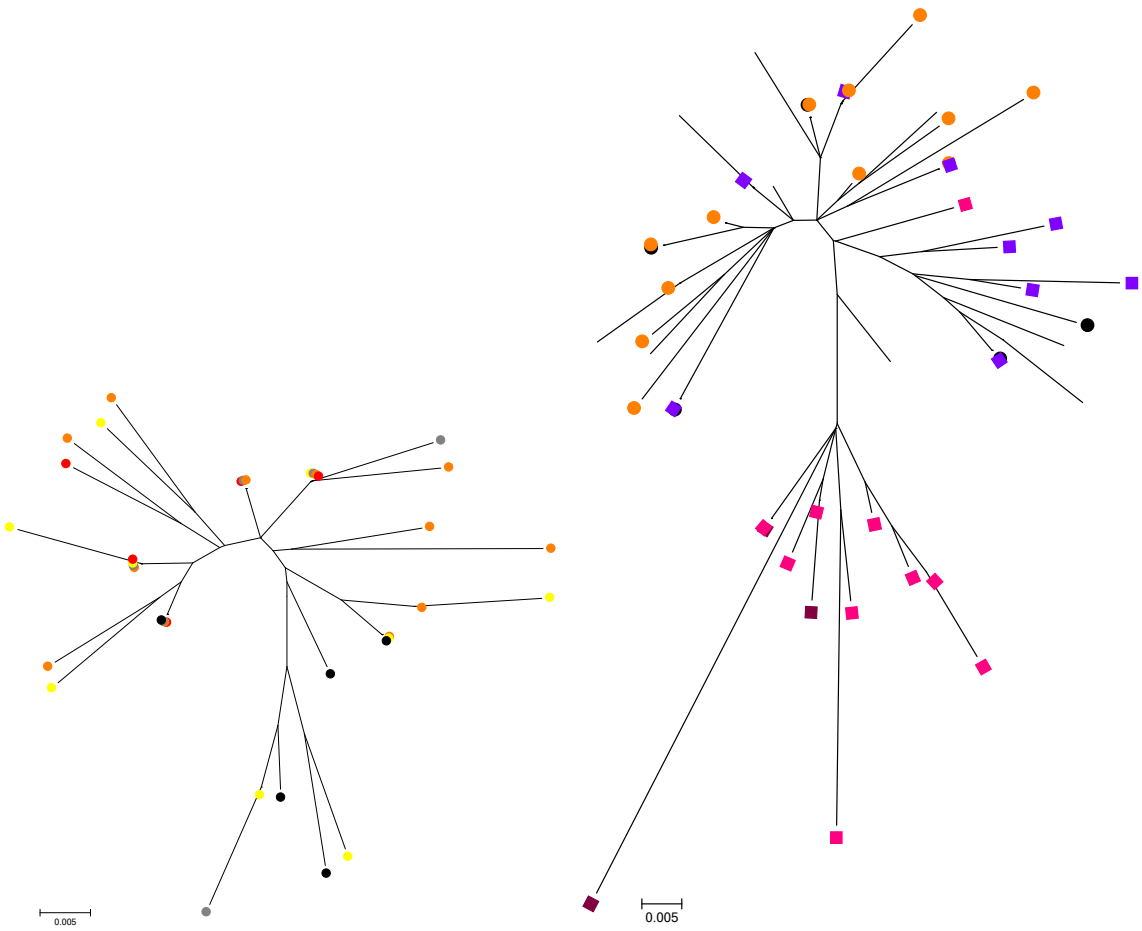
There is very little HVR1 divergence between samples.



6.3.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones. The clones demonstrate little divergence from 10 years prior to study commencement (R1).

E1 demonstrates minimal divergence throughout the study period. The least divergent HVR1 sample is the pre treatment sample suggesting that minimal divergent drift has been followed by convergent change.

6.3.1 Phylogenetic analysis



6.3.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective samples. Retrospective (R1) which was taken 10 years prior to the study onset (wine) and samples from week 16 (black) and week 0 (orange) labelled. Additional retrospective samples are labelled blue (R2 -1 years prior to study onset) and pink (5 year prior to study onset). Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation.

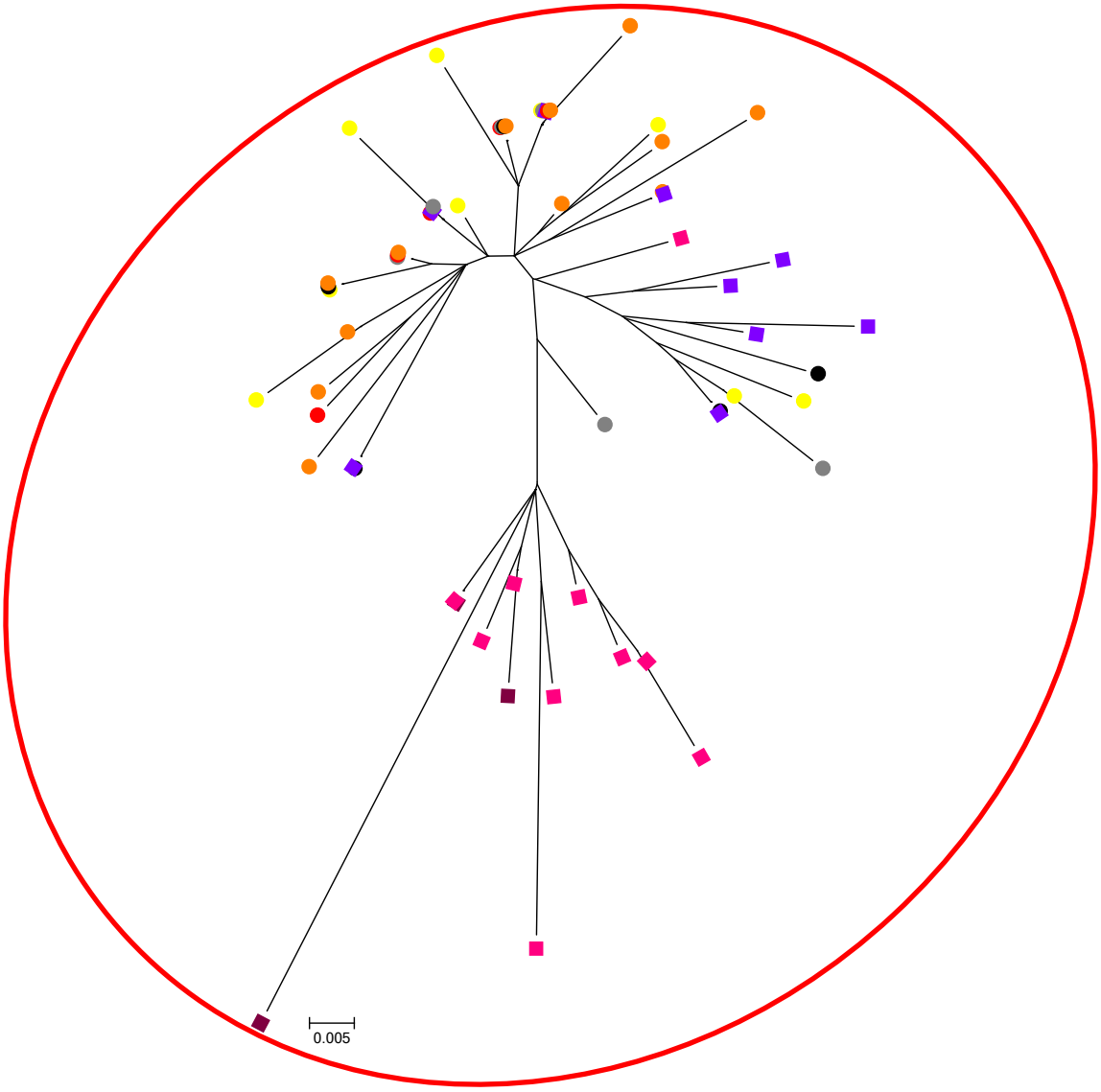
It is noticeable that the general shape of the tree has been unaffected by the inclusion of the retrospective sample.

6.3.1 Subpopulation analysis

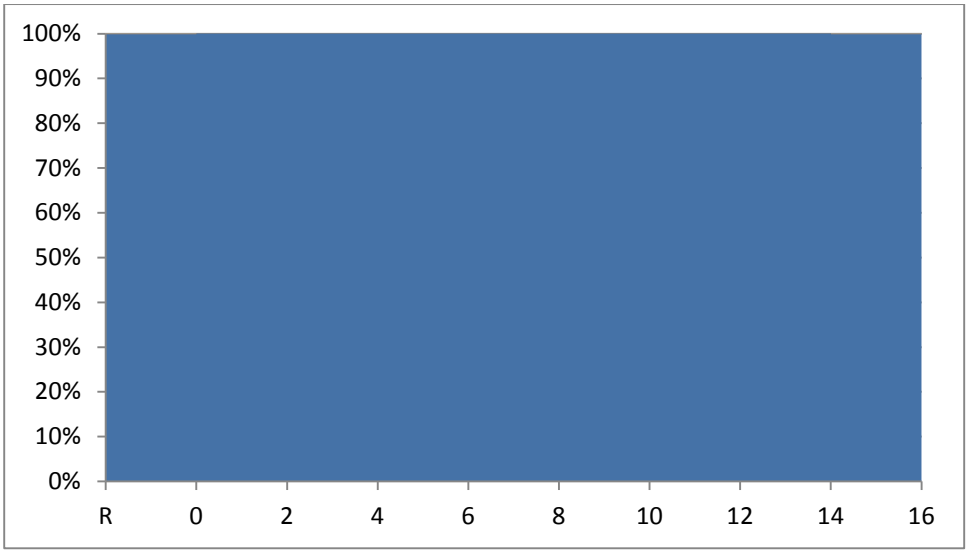
```
                          1        10        20      27
                          |--------+---------+------|
1312-0_1812-8-5_M13u      ETHITGGSQARTTRGFANLFSPGPSQN
1312-0-9_M13uni_--_1      ETHITGGSQARTTRGFANLFSPGPSQN
OA13_M13uni-21_--_8.      ETHITGGSQARTTRGFANLFSPGPSQN
1312-8-21_3007-4-9_M      ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-12_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-13_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-20_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-2_M13uni_--_1      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-7_M13uni_--_4      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-6_M13uni_--_1      ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-19_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-8-8_M13uni_--_1      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-3_M13uni_--_9      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-19_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-18_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-0-17_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-12-1_M13uni312-      ETHITGGTQARTTRGFANLFSPGPSQN
1312-12-2_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-124_M13uni_--_1      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1211_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1216_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-141_M13uni_--_1      ETHITGGTQARTTRGFANLFSPGPSQN
OB-14_M13uni-21_--_9      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1410_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1415_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1417_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1423_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
OA1_M13uni-21_--_45.      ETHITGGTQARTTRGFANLFSPGPSQN
OA23_M13uni-21_--_9.      ETHITGGTQARTTRGFANLFSPGPSQN
1312-148_M13uni_--_9      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1611_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
1312-1613_M13uni_--_      ETHITGGTQARTTRGFANLFSPGPSQN
OA11_M13uni-21_--_26      ETHITGGTQARTTRGFANLFSPGPSQN
OA1a_M13uni-21_--_45      ETHITGGTQARTTRGFANLFSPGPSQN
OA16_M13uni-21_--_21      ETHITGGTQARTTRGFANLFSPGPTQN
1312-8-10_M13uni_--_      ETHITGGAQARTTRGFANLFSPGPSQN
1312-0-4_M13uni_--_2      ETHITGGAQARTTRGFANLFSPGPSQN
1312-0-5_M13uni_--_1      ETHITGGAQARTTRGFANLFSPGPSQN
1312-8-4_M13uni_--_2      ETHITGGAQARTTRGFANLFSPGPSQN
            1312-1619     ETHITGGAQARTTRGFANLFSPGPSQN
OA10_M13uni-21_--_11      ETHITGGAQARTTRGFANLFSPGPSQN
1312-16-1_312-16-2_M      ETHITGGAQARTTRGFANLFSPGPSQN
1312-1618_M13uni_--_      ETHITGGAQARTTRGFANLFSPGPSQN
1312-8-5_M13uni_--_1      ETHITGGAQARTTRGFANLFSPGPSQN
OA8_M13uni-21_--_11.      ETHITGGAQARTTRGFANLFSPGPSQN
1312-0-1_M13uni_--_1      ETHITGGTQARTTRGFANLFSPGTSQN
O-11_M13uni-21_--_11      ETHITGGTQARTTRGFANLFSAGPSQN
OB-1_M13uni-21_--_19      ETHITGGTQARTTRGFANLFSAGPSQN
OB-15_M13uni-21_--_9      ETHITGGTQARTTRGFANLFSAGPSQN
OB-23_M13uni-21_--_9      ETHITGGTQARTTRGFANLFSAGPSQN
OB-16_M13uni-21_--_2      ETHITGGTQARTTRGFANLFSAGPSQN
OB-18_M13uni-21_--_8      ETHITGGTQARTTRGFANLFSAGPSQN
OB-21_M13uni-21_--_9      ETHITGGTQARTTRGFANLFSAGPSQN
OB-10_M13uni-21_--_1      ETHITGGAQARTTRGFANLFSAGPSQN
OA12_M13uni-21_--_25      ETHITGGTQARTTRGFANLFSPGPPQN
OB-12_M13uni-21_--_1      ETHITGGTQARTTRGFANLFSAGPPQN
O-1_M13uni-21_--_8..      ETHITGGTQARTTRGLANLFSAGPSQN
O-12_M13uni-21_--_45      QTHITGGAQARTTRGLANLFSPGPSQN
OB-20_M13uni-21_--_8      ETHITGGTQARTTRGFANLFSAEPSQN
             Consensus    ETHITGGtQARTTRGfANLFSpGPSQN
```

1

6.3.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. Encompassing 10 years of chronic infection all sequences are included in a single subpopulation. During the 10 year period under investigation, the HVR1 changes by only two amino acid substitutions.



6.3.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The retrospective samples are labelled with the same colours as 6.3.1 Fig 5. The prospective labels are: Week 16 – black, Week 14 – grey, Week 12 – red, Week 8 – yellow, Week 0 – orange.

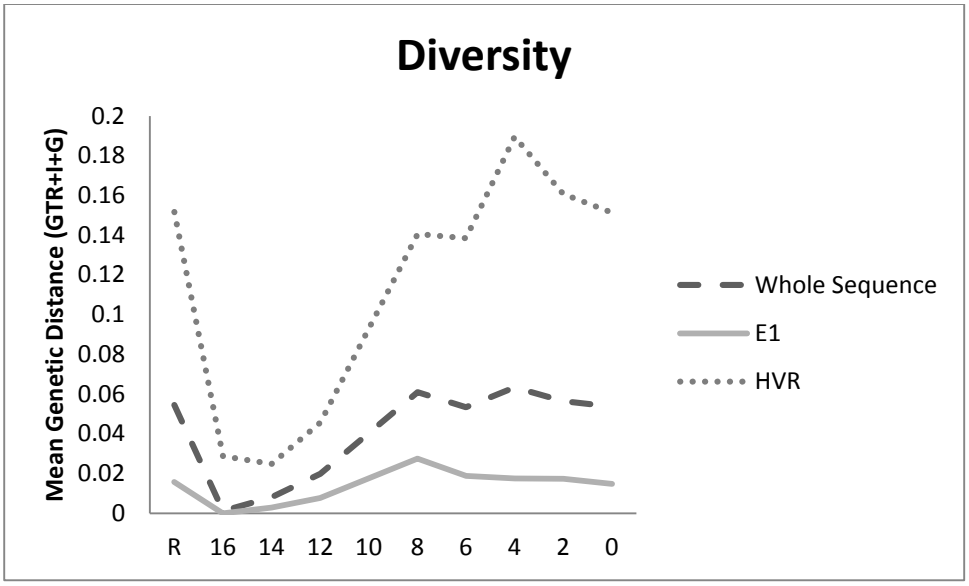


6.3.1 Fig. 8 The persistence of the single subpopulation from the retrospective sample through the study period to the pre treatment sample.

All clones generated including those from all three retrospective samples a within the same subpopulation.

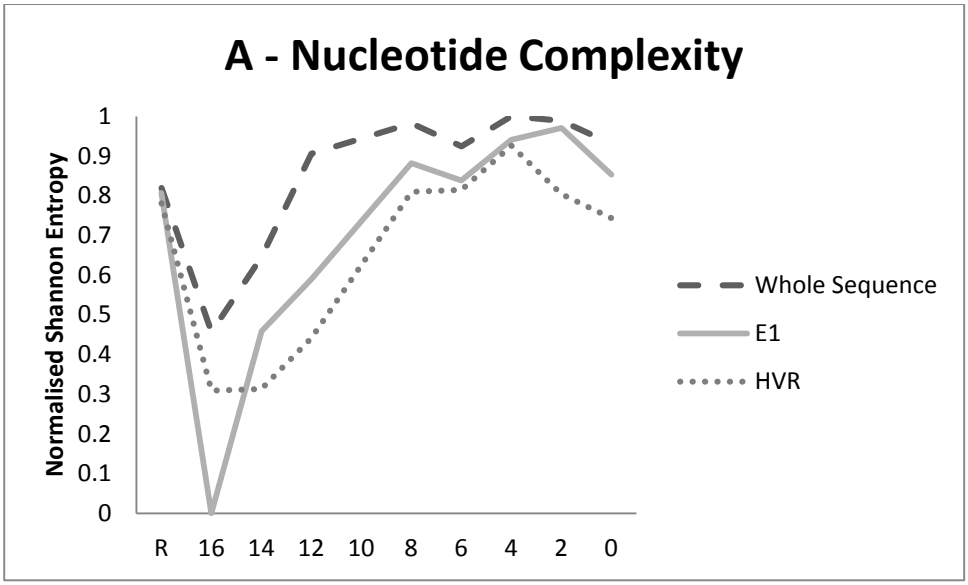
6.3.2 Subject F

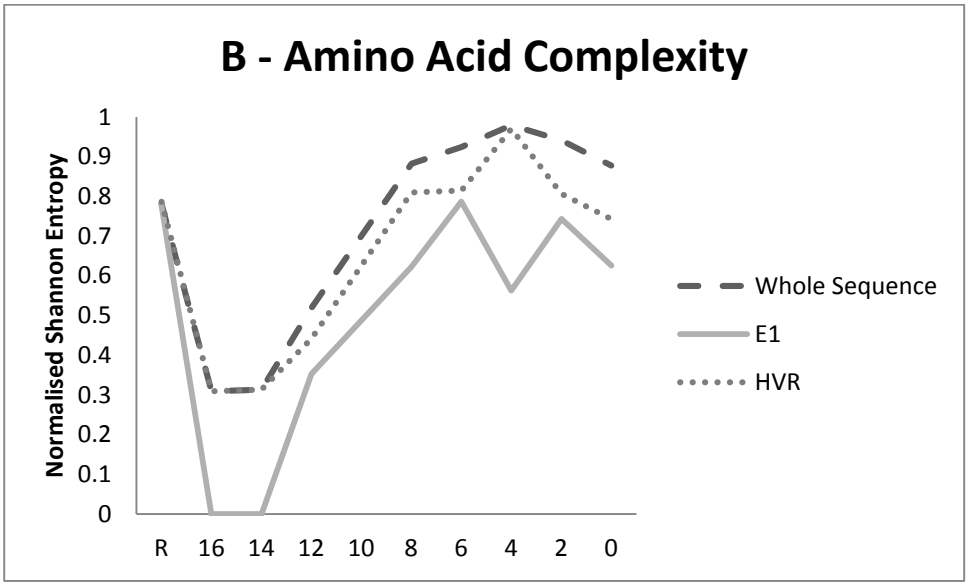
6.3.2 Diversity, Complexity, and Divergence



6.3.2 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G). The episodic high diversity is suggestive of the transient presence of co-existing multiple HVR1 subpopulations within the QS.

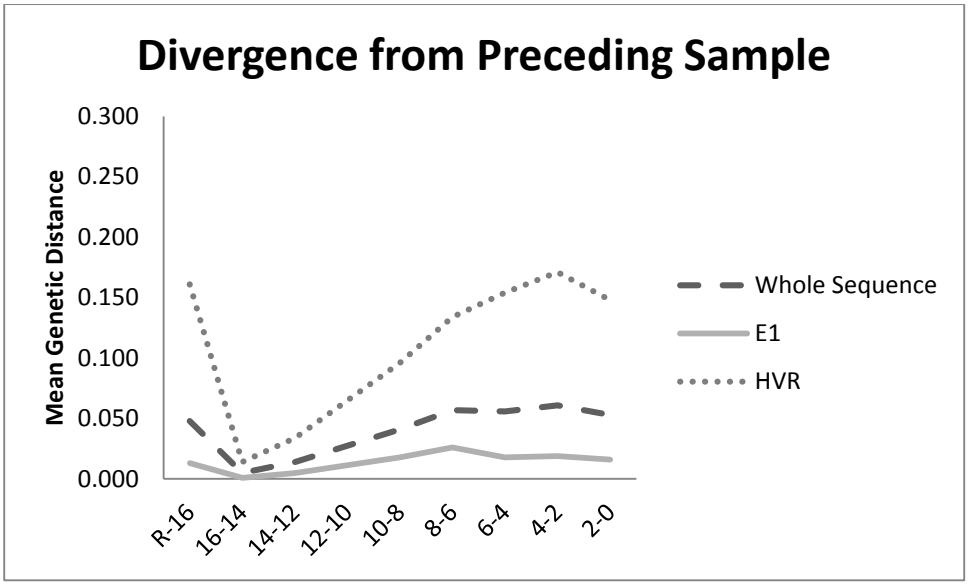
HVR1 diversity is greater than E1 diversity in all samples





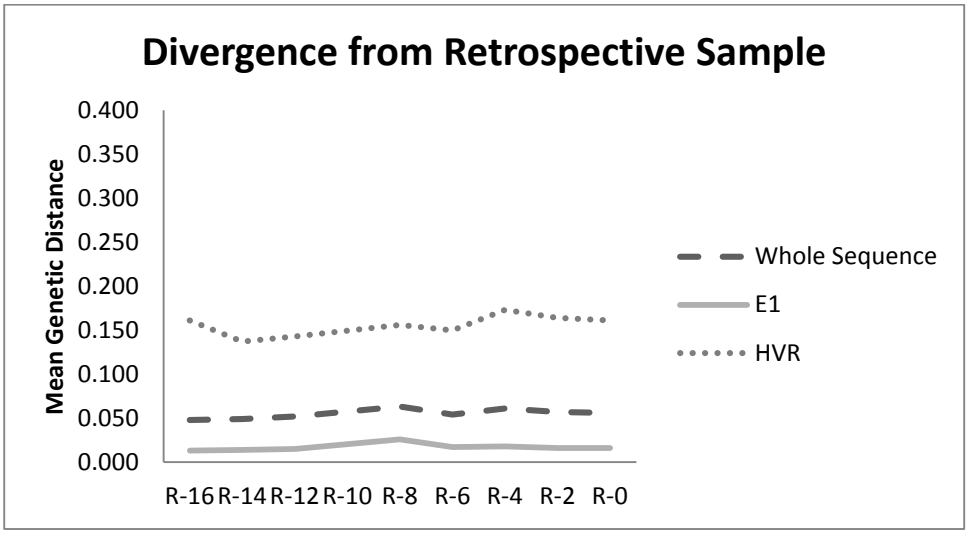
6.3.2 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased amino acid complexity relative to E1 throughout the study period.



6.3.2. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones. The retrospective sample was taken 163 days prior to the onset of the prospective study which itself lasted 112 days. The magnitude of HVR1 divergence seen between the retrospective groups of clones and that seen between Week 4 and Week 2 is similar but it must be noted from 6.3.2. Fig 1 that these samples have high diversity perhaps implying that multiple subpopulations may be distorting the true rate of divergence.

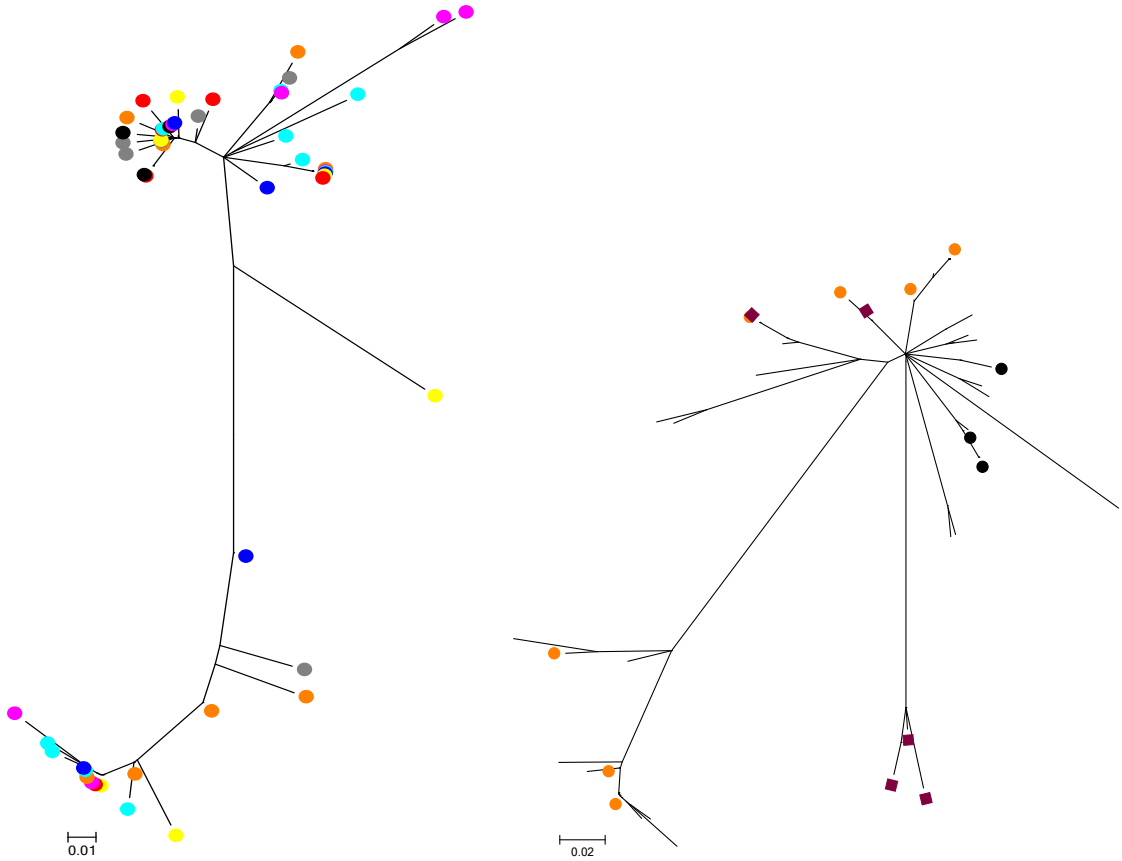
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that there is a similar magnitude of divergence.



6.3.2. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 cumulative divergence from the retrospective group is similar for all subsequent samples. Although Fig. 3 suggests significant inter sample divergence, the mean pairwise genetic distance from the retrospective sample remains constant Fig. 4 suggesting constant exploration of the sequence space at a near constant distance from the retrospective sample.

6.3.2 Phylogenetic analysis



6.3.2 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (163 days prior to Week 16 clones which are labelled with black circles. Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.

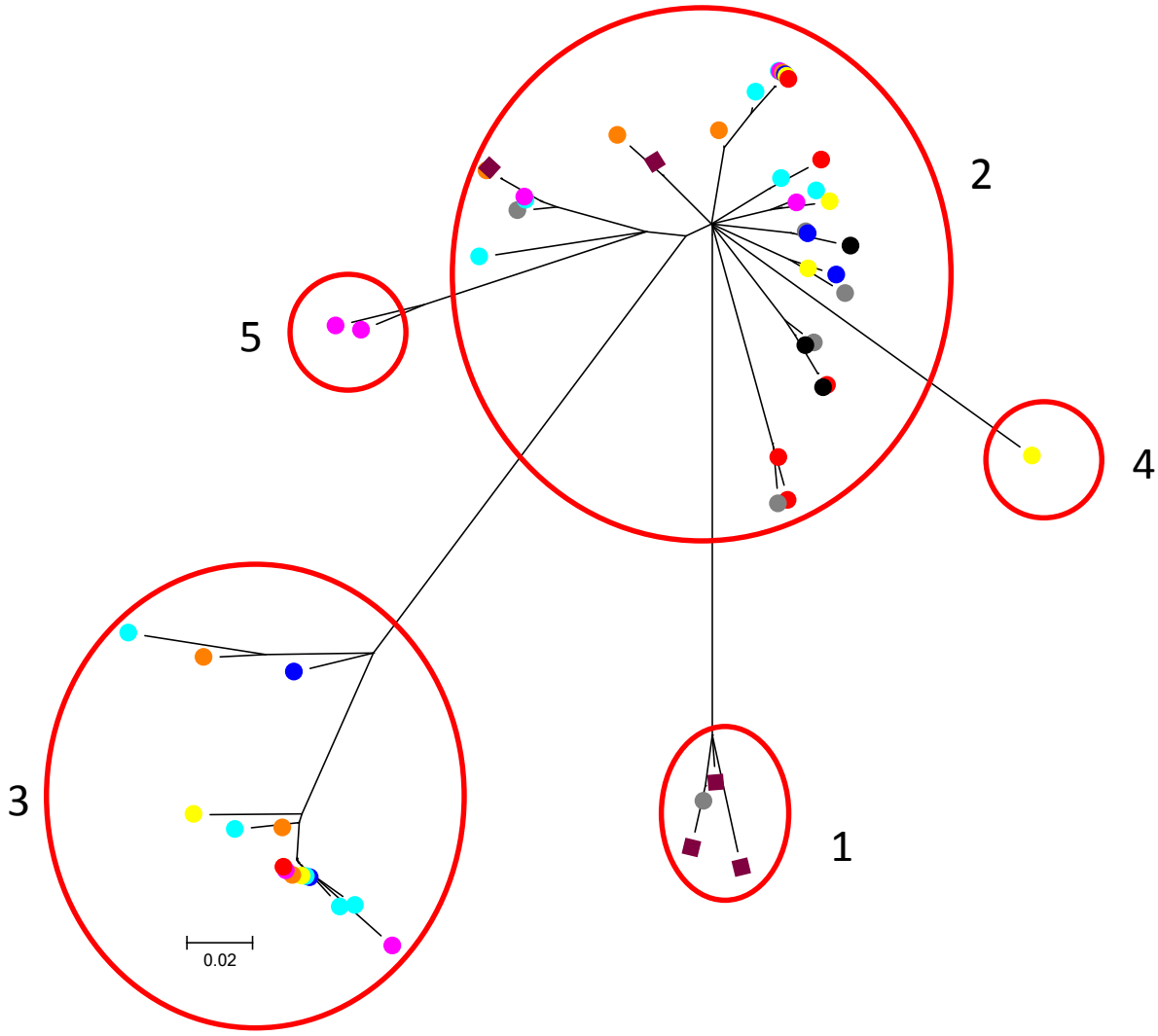
It is noticeable that the general shape of the tree has been affected by the inclusion of the retrospective sample. A number of retrospective sequences have formed a new clade which joins the majority of the sequences. This has also drawn a sample 14 sequence (grey) away from the lower sequences where it had been placed in the left tree.

6.3.2 Subpopulation analysis

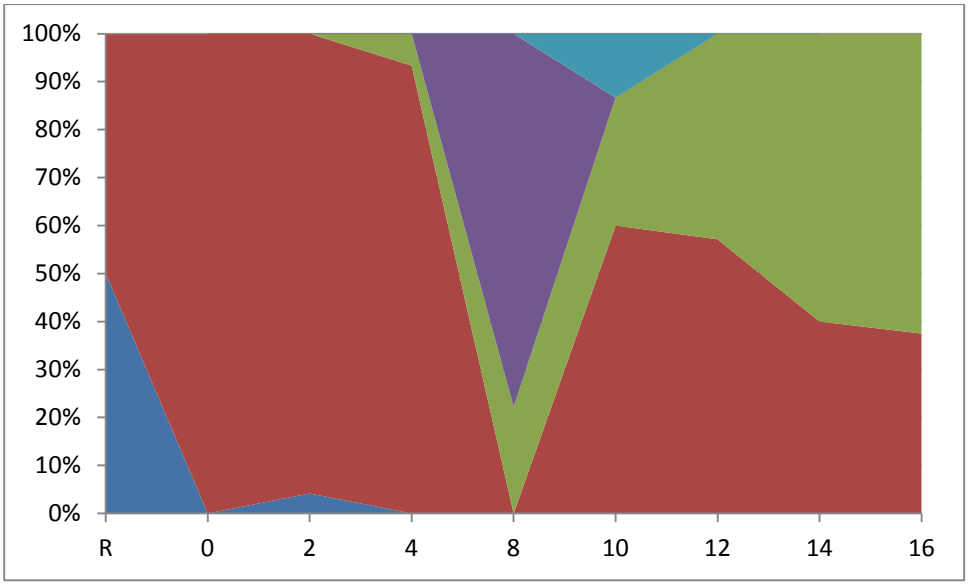
```
                                1        10        20      27
                                |--------+---------+------|
                     x0-17      GTHVTGGSAARDAFRLTNLFSVGAQQK
                      X0-8      NTHVTGGSAARDAFRLTNLFSVGAQQK
                     x0-16      NTHVTGGSAARDAFRLTNLFSVGAQQK
X-2-12_M13uni_--_9..            NTHVTGGSAARDAFRLTNLFSVGAQQK
X-2-14_M13uni_--_9..            NTHVTGGSAARDAFRLTNLFSVGAQQK
X-2-4_M13uni_--_11..            NTHVTGGSAARDAFRLTNLFSVGAQQK
X-6-2_M13uni_--_12..            NTHVTGGSAARDAFRLTNLFSVGAQQK      3
      1505810_M13uni            NTHVTGGSAARDAFRLTNLFSVGAQQK
     15051215_M13uni            NTHVTGGSAARDAFRLTNLFSVGAQQK
1505-4-9_M13uni_--_2            NTHVTGGSAARDAFRLTNLFSVGAQQK
X-2-11_M13uni_--_9..            NTHVTGGSAARHAFRLTNLFSVGAQQK
1505-4-1_M13uni_--_5            NTHITGGSAARDAIRLTNLFSVGAQQK
X-6-1_M13uni_--_14..            GTHVTGGSAARGAFRLTNLFSVGAQQK
X-10_M13uni-21_--_10            GTHVTGGAAARDAFRLTNLFSAGAQQK
X-2-18_M13uni_--_8..            GTHVTGGAAARDAFGLANLFSVGAQQK
       150587_M13uni            NTHVTGGAAARDAFGLANLFSVGAQQK
X-14-12_M13uni_--_15            GTHVTGGAAARDAFGLTSLFSAGAQQK
    X-3_M13uni-21_--            GTHVTGGAAARDAFGLTSLFSAGAQQK      1
X-6_M13uni-21_--_91.            NTHVTGGAAARDAFGLTSLFSAGAQQK
                      X0-7      GTYVTGGSAARAASTFAGLFTPGAQQK
X-18_M13uni-21_--_19            GTYVTGGSAARAASTFAGLFTPGAQQK
X-2-7_M13uni_--_20..            GTHVTGGSAARAASTFAGLFTPGAQQK
X-14-21_M13uni_--_14            GTHVTGGSAARAASTFAGLFTPGAQQK
1505-4-18_M13uni_--_            GTHVTGGSAARAASTFAGLFTPGAQQK
                      X0-1      GTHVTGGSAARGASTLAGLFTPGAQQK
                      X0-2      GTHVTGGSAARGASTLAGLFTPGAQQK
X-2-21_M13uni_--_51.            GTHVTGGSAARGASTLAGLFTPGAQQK
        150666-16-19            GTHVTGGSAARGASTLAGLFTPGAQQK
150566-16-150566-16-            GTHVTGGSAARGASTLAGLFTPGAQQK
X-14-1_M13uni_--_26.            GTHVTGGSAARGASTLAGLFTPGAQQK
X-14-14_M13uni_--_22            GTHVTGGSAARGASTLAGLFTPGAQQK
X-14-6_M13uni_--_15.            GTHVTGGSAARGASTLAGLFTPGAQQK
X-14-22_M13uni_--_11            GTHVTGGSAARGASTLAGLFTPGAQQK
1505-4-12_M13uni_--_            GTHVTGGSAARGASTLAGLFTPGAQQK
1505123_M13uni505122            GTHVTGGSAARGASTLAGLFTPGAQQK
     15051211_M13uni            GTHVTGGSAARGASTLAGLFTPGAQQK      2
      1505128_M13uni            GTHVTGGSAARGASTLAGLFTPGAQQK
      1505124_M13uni            GTHVTGGSAARGASTLAGLFTPGAQQK
X-1_M13uni-21_--_22.            GTHVTGGSAARGASTLAGLFTPGAQQK
X-6-16_M13uni_--_10.            GTHVTGGSAARGASTLAGLFTPGAQQK
150589_M13uni505813_            GTHVTGGSAARGASTLAGLFTPGAQQK
      1505815_M13uni            GTHVTGGSAARGASTLAGLFTPGAQQK
         150566-16-4            GTHVTGGSAVRGASTLAGLFTPGAQQK
X-6-23_M13uni_--_22.            GTHVTGGSAARGASALAGLFTPGAQQK
X-2-3_M13uni_--_21..            GTHVTGGSAAHGASTLAGLFTPGAQQK
                      X0-6      GTHVTGGSAAHGASTLAGLFTRGAQQK
X-2-22_M13uni_--_11.            GTHVTGGSAAHGASTLAGLFTRGAQQK
1505-4-19_M13uni_--_            GTHVTGGSAAHGASTLAGLFTRGAQQK
X-6-24_M13uni_--_22.            GTHVTGGSAAHGASTLAGLFTRGAQQK
       150585_M13uni            GTHVTGGSAAHGASTLAGLFTRGAQQK
     15051212_M13uni            GTHVTGGSAAHGASTLAGLFTRGAQQK
X-2-1_M13uni_--_8..9            GTHVTGGSAARGASTLAGLFTRGAQQK
X-2-10_M13uni_--_18.            NTHVTGGSAARGASTLAGLFTRGAQQR
       150582_M13uni            GTHVTGGSAARAASGLAGLFSPGAQQK      4
1505-4-20_M13uni_--_            DTHVTGGSVARGASAFAGLFSQGAQQK
1505-4-3_M13uni_--_3            DTHVTGGSVARAASSFAGLFSQGAQQK
           Consensus            gTHVTGGsaAR.As.lagLFs.GAQQK      5
```

6.3.2 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (5 in total) are designated by

red integers. The numbering of subpopulations was done in accordance with the temporal appearance of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



6.3.2 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Multiple co-existing subpopulations are identified with the emergence and elimination of novel subpopulations during the prospective study period. There appears to be temporal change in the dominant subpopulation during the study.

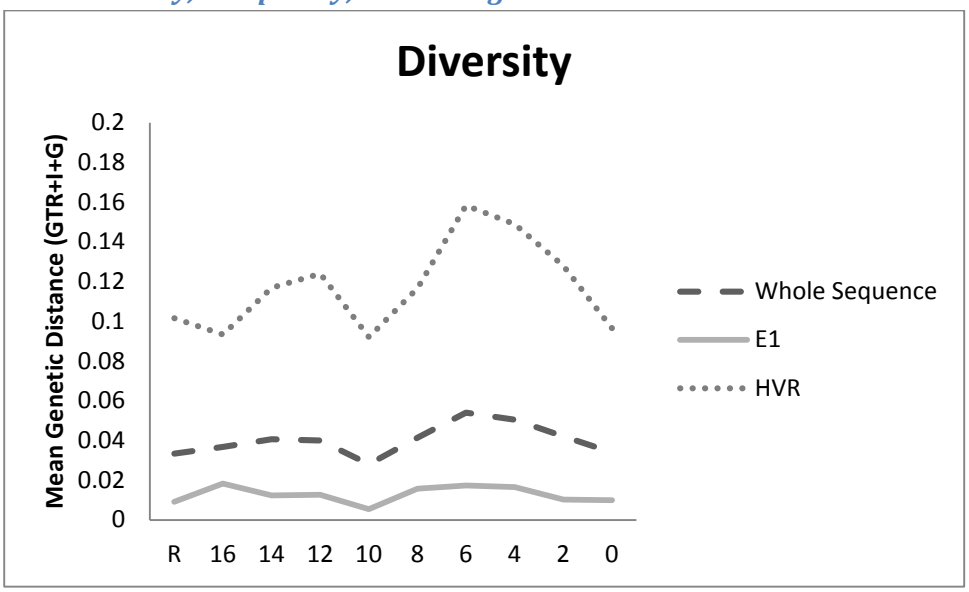


6.3.2 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subpopulation 1 (blue) co dominant in the retrospective sample and a minor component of week 14 sequences is subsequently completely eliminated from the HVR1 QS. Subpopulation 3 (green) emerges in the pre treatment samples and increases in prevalence and makes up 60% of the QS in the pre treatment sample. This feature suggests a time order phylogeny.

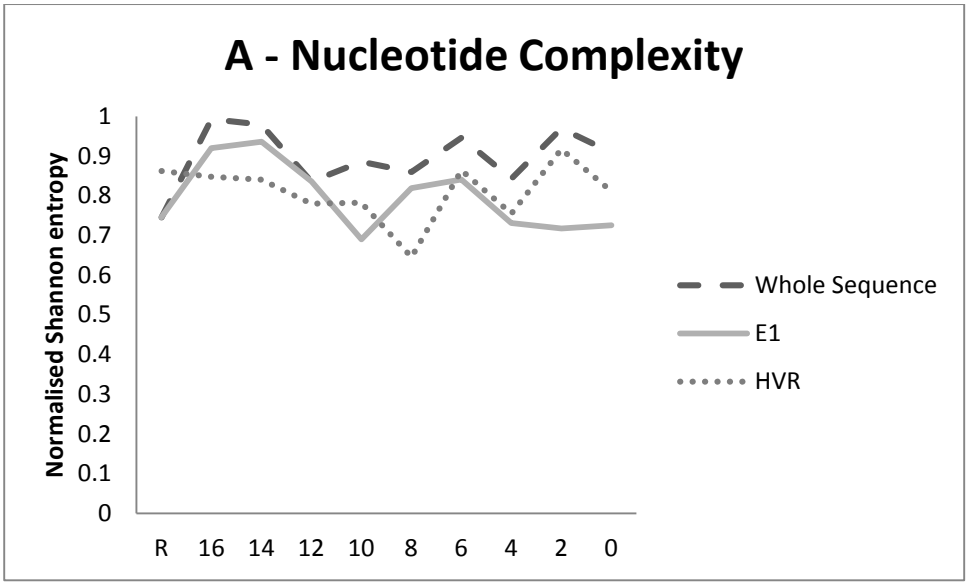
6.3.3 Subject H

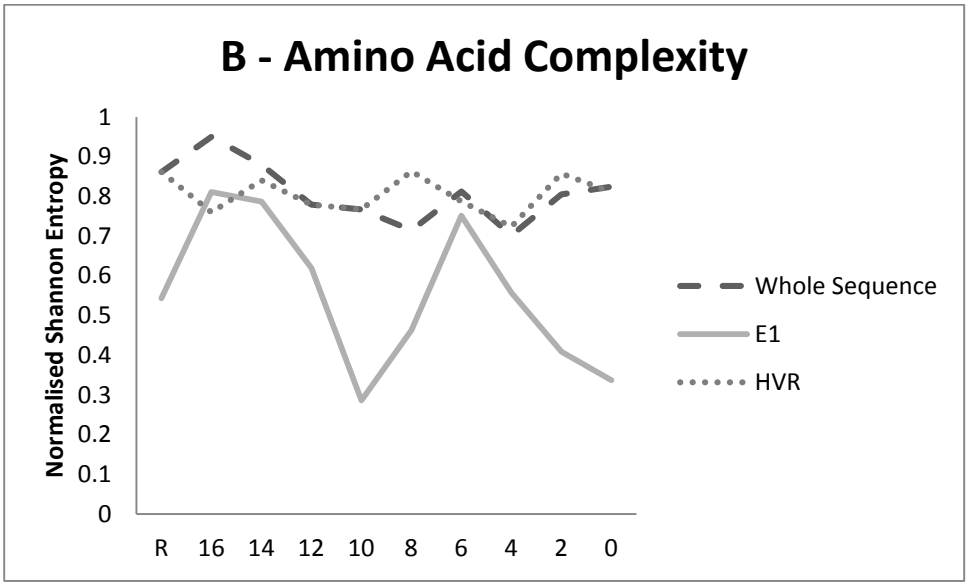
6.3.3 Diversity, Complexity, and Divergence



6.3.3 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G). High diversity in the HVR clones reflects remotely related groups of clones and suggests the presence of multiple HVR1 subpopulations.

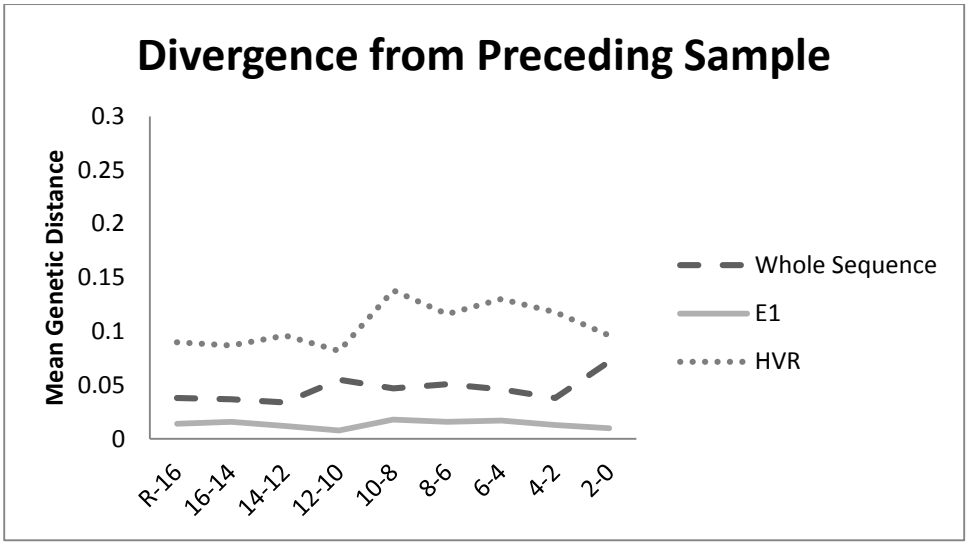
HVR1 diversity is greater than E1 diversity.





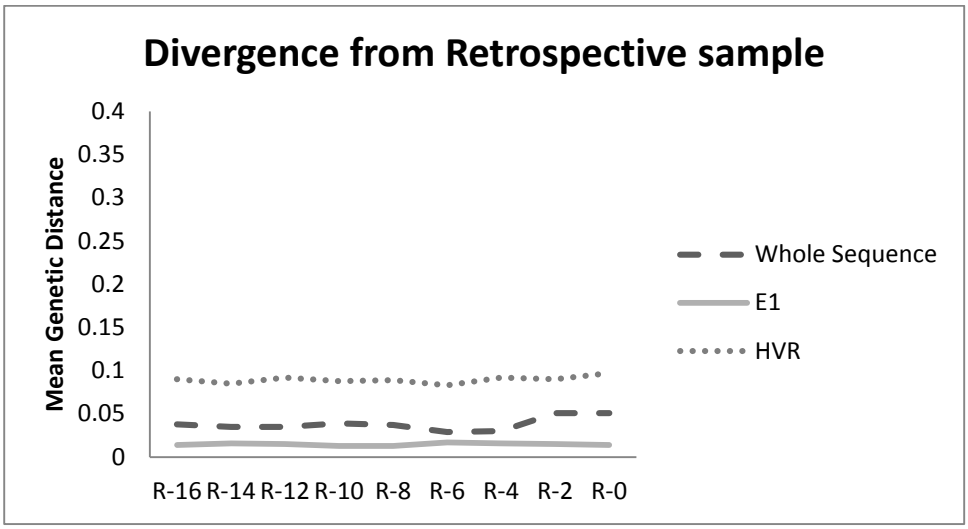
6.3.3 Fig 7. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased complexity relative to E1 throughout the study period.



6.3.3. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

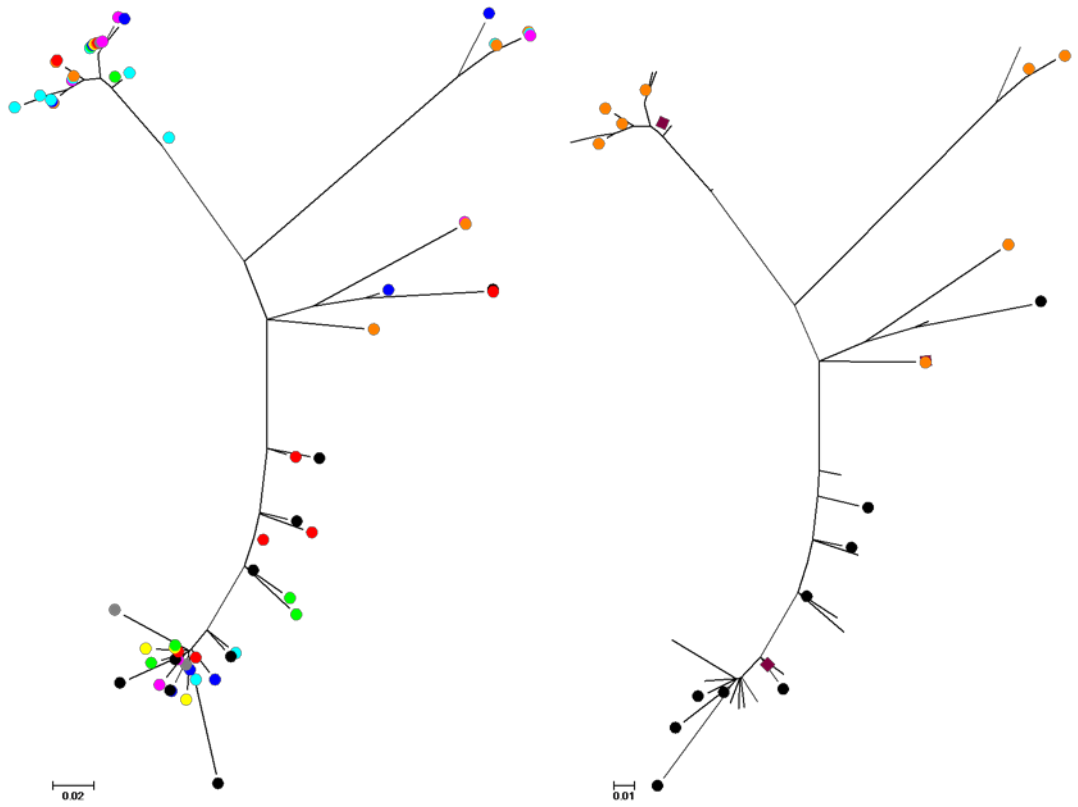
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that there is a similar magnitude of divergence.



6.3.3. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

There is minimal divergence throughout the study interval and including the retrospective sample.

6.3.3 Phylogenetic analysis



6.3.3 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample. Retrospective (wine) (176 days prior to prospective study) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.

The phylogenetic tree is unaffected by the inclusion of the retrospective sample.

Subject H had in the 16 weeks prior to commencing treatment demonstrated characteristics suggestive of a time ordered phylogeny with complete replacement of the initially dominant HVR1 subpopulation with a subpopulation not identified in the sample taken 16 weeks prior to commencing treatment. However, the retrospective sample includes a sequence which suggests that there may be a sustained mixed lineage infection rather than a time order phylogeny.

6.3.3 Subpopulation analysis

```
                          1        10        20      27
                          |--------+---------+------|
B-1_M13uni-21_--_47.      TTYATGASAGRSAQGLTSLFSVGPRQN
251708_M13uni_--_38.      TTYATGASAGRSAQGLTSLFSVGPRQN
B-17_M13uni-21_--_25      TTYVTGASTGRGAKGLASLFTLGPQQN
251-16-8_M13uni_--_1      TTYVTGASTGRGAKGLASLFTLGPQQN
B-2-8_M13uni_--_16..      TTYVTGASTGRGARGLASLFTLGPQQN
B-4-19_M13uni_--_22.      TTYVTGASTGRGARGLASLFTLGPQQN
B-6-4_M13uni_--_20..      TTYVTGASTGRGARGLASLFTLGPQQN
251-12-10_M13uni_--_      TTYVTGASTGRGARGLASLFTLGPQQN
25170-85_M13uni_--_1      TTYVTGASTGRGARGLASLFTLGPQQN
B-10-21_M13uni_--_26      TTYVTGASTGRGARGLASLFTLGPQQN
B-14-1_M13uni_--_21.      TTYVTGASTGRGARGLASLFTLGPQQN
251-16-19_M13uni_--_      TTYVTGASTGRGARGLASLFTLGPQQN
B-10-14_M13uni_--_11      TTYVTGASTGRGARGLASLFTLGPKQN
251-12-4_M13uni_--_8      TTYVTGASTGRGARGLASLFTLGPKQN
251-16-6_M13uni_--_3      ATYVTGASTGRGARGLASLFTLGPQQN
B-4-11_M13uni_--_13.      TAYVTGASTGRGARGLASLFTLGPQQN
251-12-15_M13uni_--_      TTYVTGASAGRGARGLASLFTLGPKQN
251-16-13_M13uni_--_      TTYVTGASTGRGAQGLTSLFTLGPQQN
251-16-4_M13uni_--_7      TTYVTGASTGRGARGLASLFTLGPHHN
B-14-10_M13uni_--_20      TTYVTGASTGKGARGLASLFTSGPQQN
251-12-3_M13uni_--_1      TTYVTGASTGRGARGLTGLFTLGPKQN
B-10-2_M13uni_--_10.      TTYVTGASTGRGARGLTSLFAVGPRQN
B-4-17_M13uni_--_19.      STRVTGANVGRGAQGLTGLFNLGPRQN
251-0-10_M13uni_--_1      STRVTGANVGRGAQGLTGLFNLGPRQN
B-6-10_M13uni_--_19.      ATYTTGASVGRGAQGLTGLFNLGPRQN
251-16-17_M13uni_--_      ATYTTGASMGRGAKGLTGLFNLGPRQN
251-12-9_M13uni_--_2      ATYTTGASMGRGAKGLTGLFNLGPRQN
B-2-13_M13uni_--_13.      TTYVTGGSVGKGVQGISSLWSSGPKQN
B-4-18_M13uni_--_17.      TTYVTGGSVGKGVQGISSLWSSGPKQN
251-0-14_M13uni_--_1      TTYVTGGSVGKGVQGISSLWSSGPKQN
B-2-5_M13uni_--_16..      TTYVTGGSVGRGVQGISSLWSSGPKQN
251-0-8_M13uni_--_15      TTYVTGGSVGRGVQGISSLWSSGPKQN
B-6-20_M13uni_--_15.      TTYVTGGSVGKGIQGLSSLWSSGPKQN
B-10_M13uni-21_--_9.      ATYVTGGSVASSAQRLTSLFTVGPRQN
B-10-3_M13uni_--_11.      ATYVTGGSVASSAQRLTSLFTVGPRQN
B-2-11_M13uni_--_14.      ATYVTGGSVASSAQRLTSLFTAGPRQN
B-4-13_M13uni_--_14.      ATYVTGGSVASSAQRLTSLFTAGPRQN
B-6-21_M13uni_--_18.      ATYVTGGSVASSAQRLTSLFTAGPRQN
251-0-1_M13uni_--_14      ATYVTGGSVASSAQRLTSLFTAGPRQN
B-2-12_M13uni_--_14.      ATYVTGGSVASSAQRLTSLFTLGPQQN
B-2-10_M13uni_--_17.      ATYVTGGSVASSAQRLTSLFAVGPRQN
B-4-1_M13uni_--_21..      ATYVTGGSVASSAQRLTSLFAVGPRQN
B-10-5_M13uni_--_11.      ATYVTGGSVASSAQRLTSLFAVGPRQN
251-12-16_M13uni_--_      ATYVTGGSVASSAQRLTSLFAVGPRQN
251702_M13uni_--_11.      ATYVTGGSVASSAQRLTSLFAVGPRQN
25170-82_M13uni_--_1      ATYVTGGSVASSAQRLTSLFAVGPRQN
B-6-14_M13uni_--_13.      ATYVTGGSVASSAQRLTSLFAVGPRQN
B-14-12_M13uni_--_13      ATYVTGGSVASSAQRLTSLFAVGPRQN
251-0-6_M13uni_--_18      ATYVTGGSVASSAQRLTSLFAAGPRQN
251-12-8_M13uni_--_3      ATYVTGGSVASSAQRLTSLFAAGPRQN
           Consensus      .TYvTGaSvgrgAqgLtsLF..GPrQN
```

2

1

5

6

7

8

9

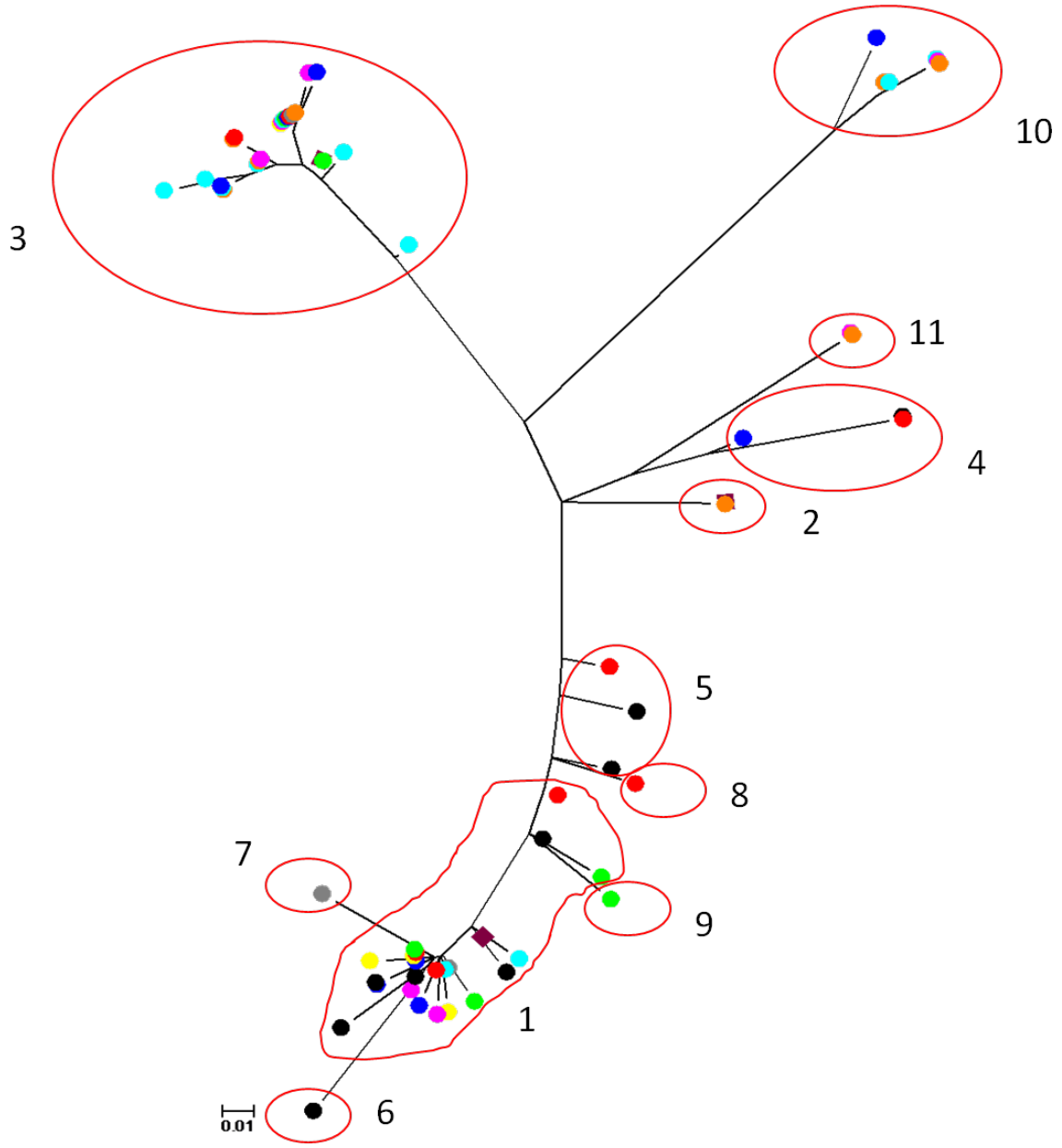
11

4

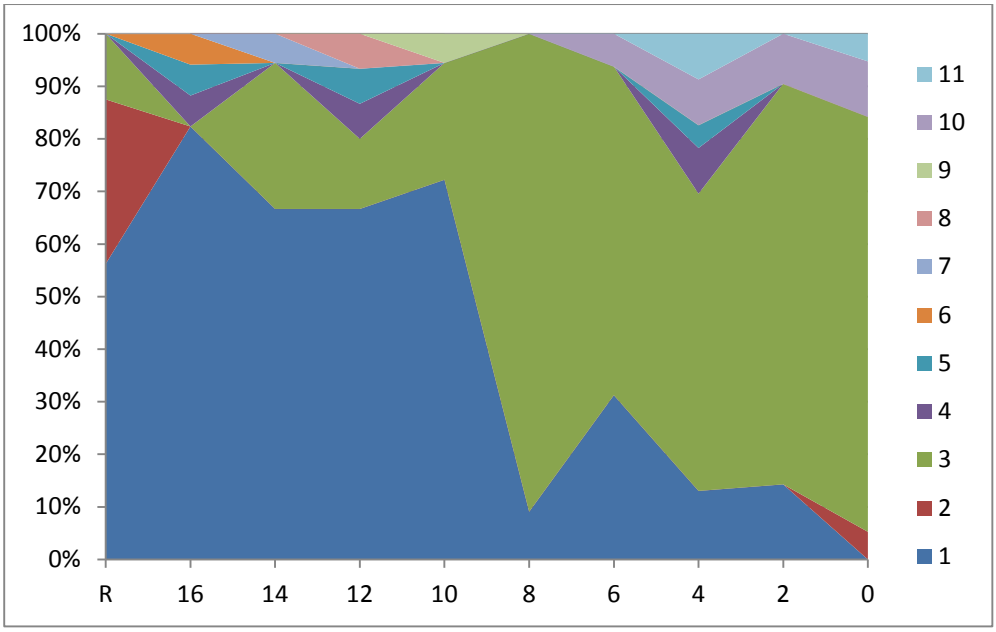
10

3

6.3.3 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (11 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



6.3.3 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.



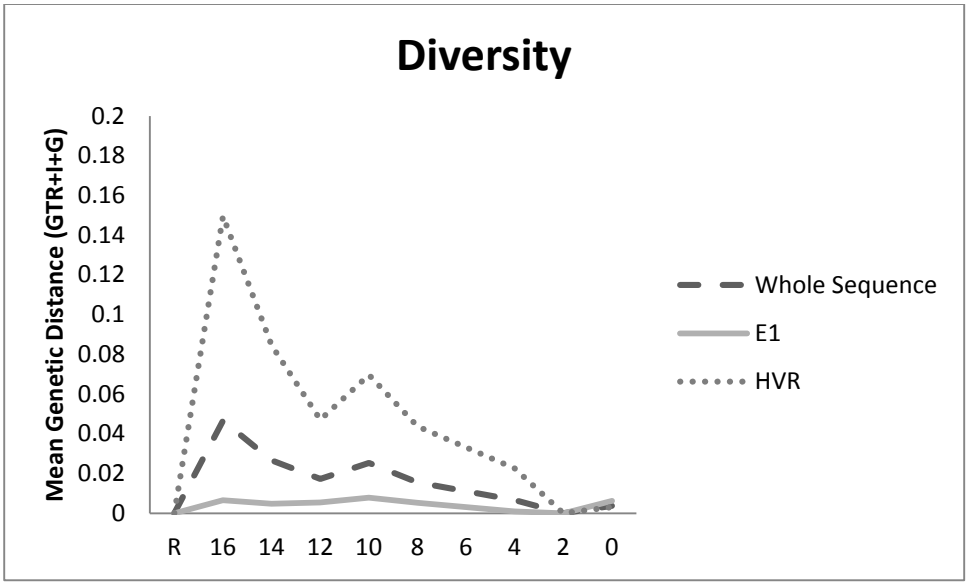
6.3.3 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

The initially dominant subpopulation makes up an ever diminishing proportion of the quasispecies through the study period. Coincident with this is the ever increasing prevalence within the clones generated of one of the minor subpopulations (green) from the retrospective sample. This subpopulation was not present among the clones generated at timepoint 16 and in prior to performing retrospective analysis of the HVR1 QS we had concluded that this subpopulation arose during the study period. The retrospective samples suggests that these multiple subpopulations may co exist in the QS over prolonged periods prior to the gradual displacement of the previous dominant subpopulation either as a result of competitive exclusion or the emergence of an immune mediated clearance of the previous subpopulation.

When examined together Figures 7 and 8 highlight the importance of subpopulations 1 and 3 who together comprise a majority of the sequences generated for each sample. It is also interesting to note the emergence of subpopulation 10 towards the end of the study period. This group of sequences is phylogenetically remote from all previous sequences and may represent the exploration of a new remote fitness optima within the sequence space.

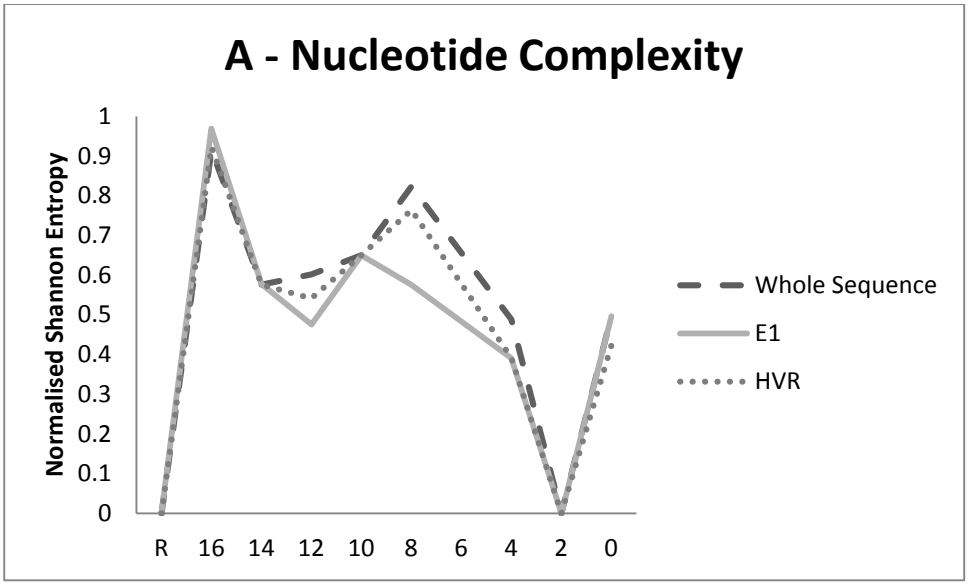
6.3.4 Subject Q

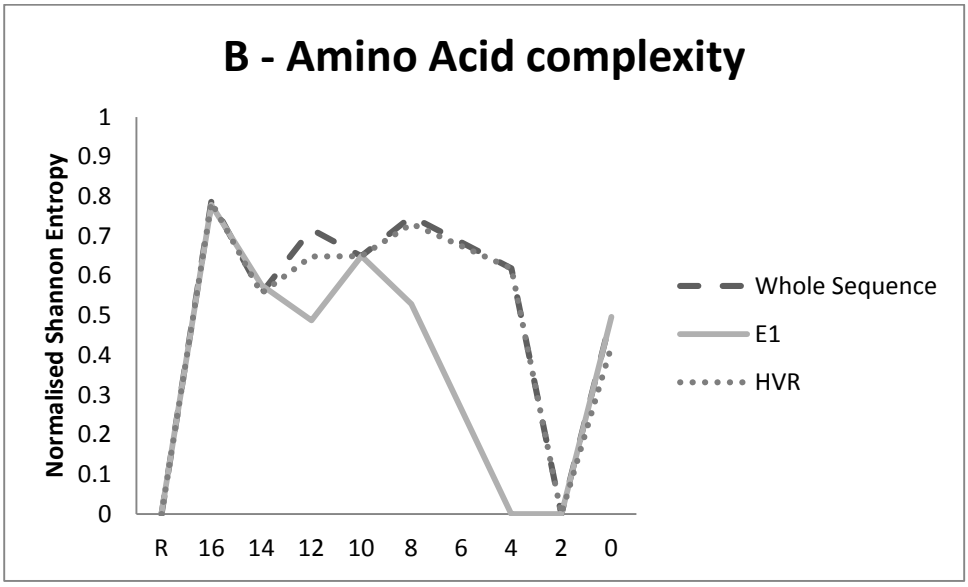
6.3.4 Diversity, Complexity, and Divergence



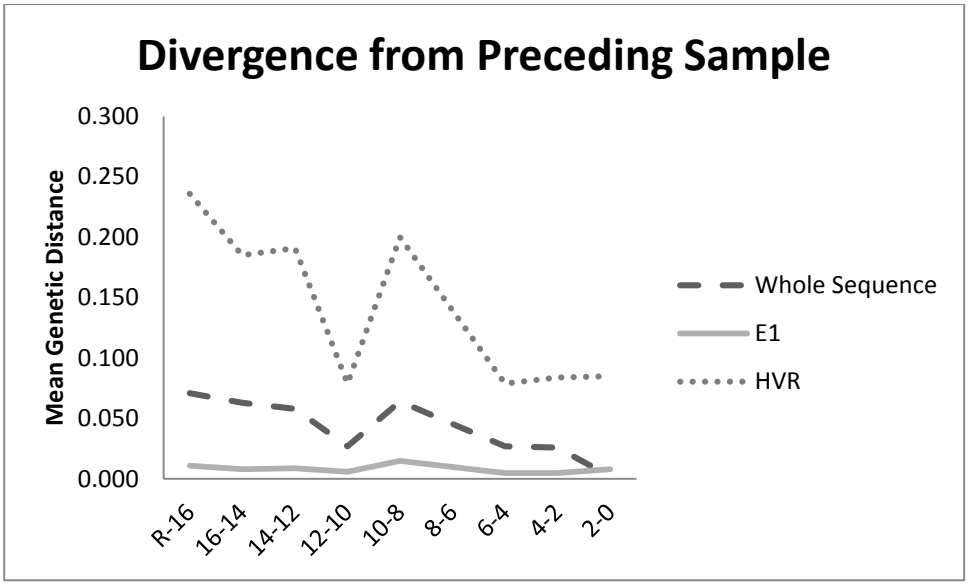
6.3.4 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

HVR1 diversity is greater than E1 diversity with the exception of the retrospective sample and the two pre treatment samples where a dramatic homogenisation of the HVR1 QS was demonstrated.



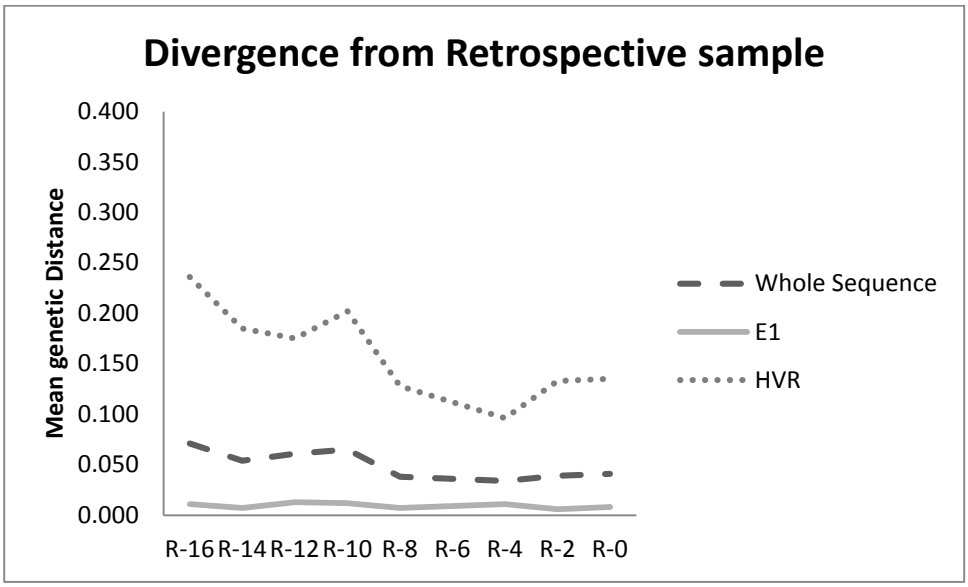


6.3.4 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.



6.3.4. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

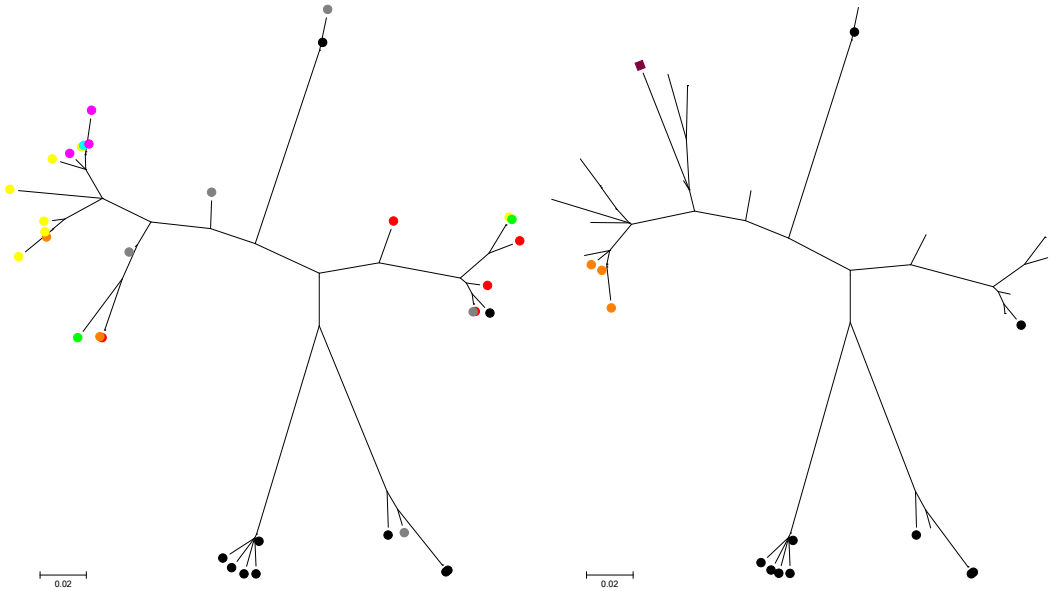
Maximal divergence occurs between the retrospective sample (378 days prior to commencement of prospective study) and the first study sample but divergence continues throughout the 16 weeks prior to commencing treatment. There is little E1 divergence.



6.3.4. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group is maximal at the sample 16 weeks pre treatment.

6.3.4 Phylogenetic analysis



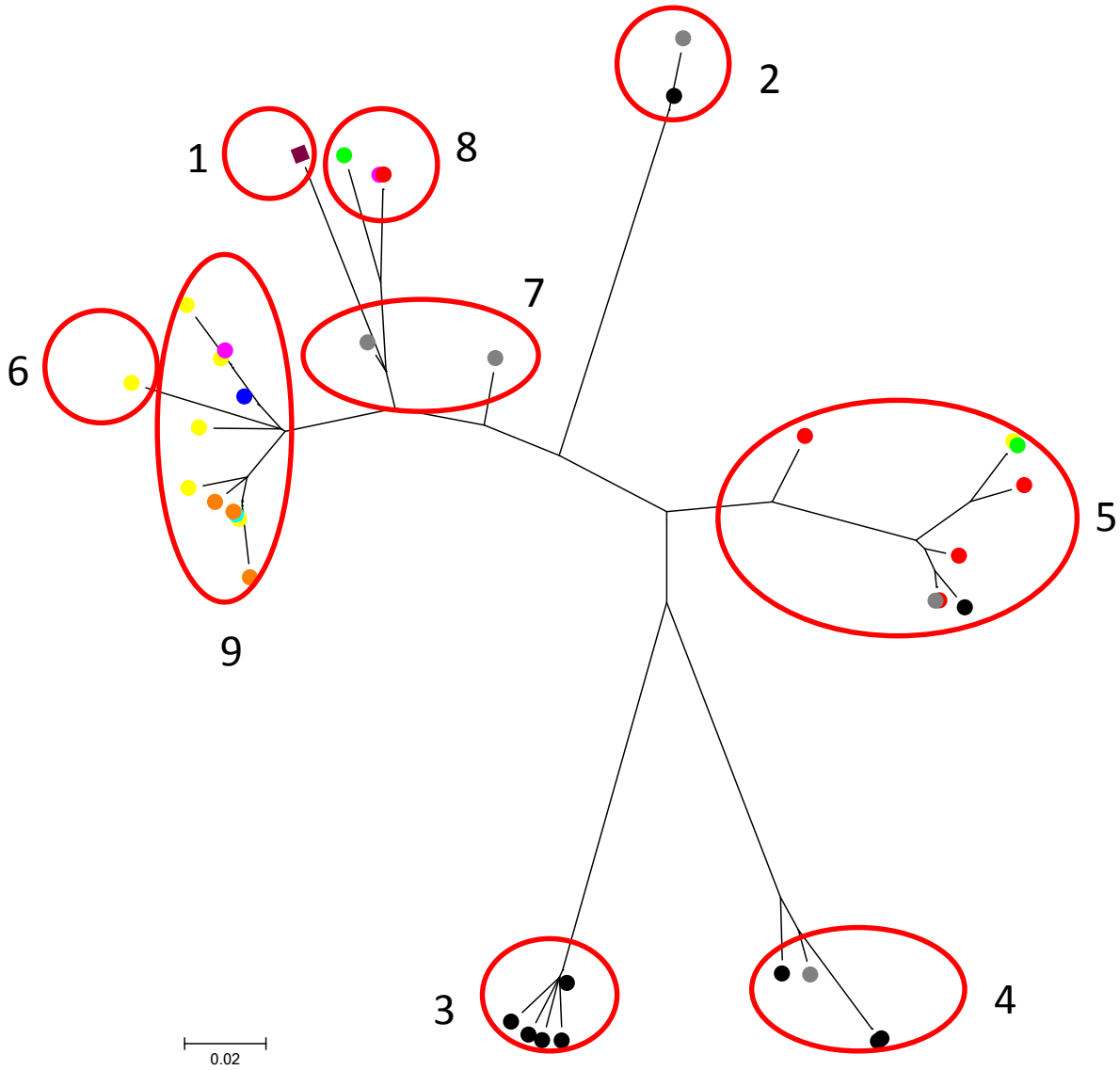
6.3.4 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (378 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.

The phylogenetic tree has been unaffected by the inclusion of the retrospective sample. The retrospective samples are situated close to a week 14 sequence. This tree illustrates why the divergence figures suggest convergence from week 16 onwards with subsequent samples more closely related to the retrospective sequences. This tree illustrates clearly a time order phylogeny with sequential samples mapping to different clades in the tree.

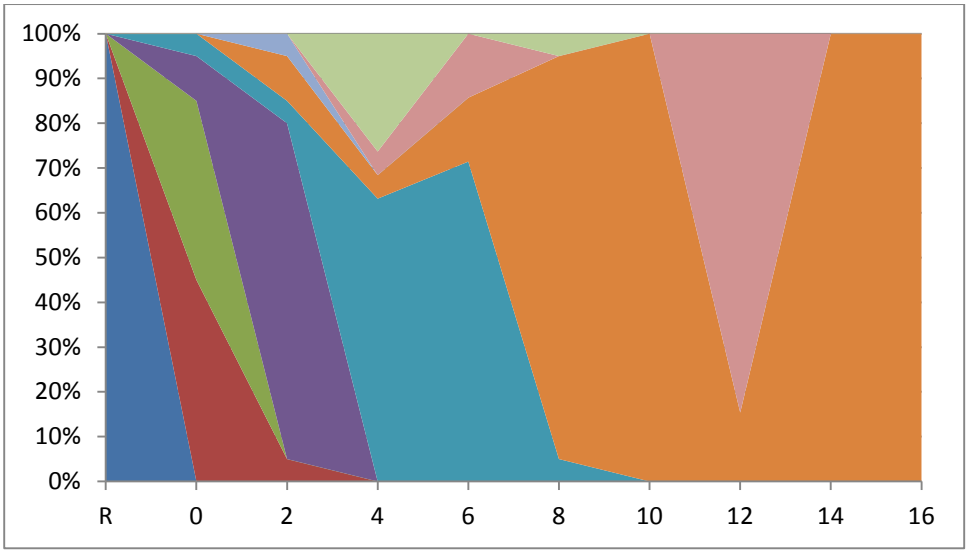
6.3.4 Subpopulation analysis

```
                        1        10         20      27
                        |--------+---------+------|
10108682_M13uni01086    ETTISGGAVARGVHSFTSLFKSGARQN  8
        101086-12-13    ETTISGGAVARGVHSFTSLFKSGARQN
M-10-22_M13uni_--_19    ETTISGGAVARGVHSFTSLFSTGARQN
  1010-141_M13uni_--    ETTISGGAAARGVHSFTSLFTPGASQK
1010-1411_M13uni_--_    ETTISGGAAARGVHSFTSLFTPGARQN
1010-213_M13uni_--_3    ETTISGGAAARGAYSLTNLFTPGARQN  7
10108677_M13uni01086    ETTISGGAAARGAYSLTNLFTPGARQN
     10108678_M13uni    ETTISGGAAARGAYSLTNLFTPGARQN
101086-8-801086-8-10    ETTISGGAAARGAYSLTNLFTPGARQN
     10108666_M13uni    ETTITGGAAARGAYSLTNLFTPGARQN
 101086-8-301086-8-5    ETTISGGAAARGAYSLTSLFTSGARQN  9
          101086-8-13   ETTISGGAAARGAYSLTNLFTSGARQN
101086-8-101086-8-20    ETTISGGAAARSAYSLTSLFTSGARQN
           101086-8-6   ETTISGGAAARSAYSLTSLFTSGARQN
10108693_M13uni01086    ETTISGGAAARSAYSLTSLFTSGARQN
M-6-10_M13uni_--_8..    ETTISGGAAARSAYSLTSLFTPGARQN
          101086-8-12   ETTITGGAAARGAYSLTRLFASGARQN  6
M-1_M13uni-21_--_22.    ETIISGGTAAYGVRSFTSLFTSGARQN
          101086-8-14   ETTITGGTVARGVSTITRLFAPGASQK  1
M-10-10_M13uni_--_21    ETTITGGTVARGVSTITRLFAPGASQK
        101086-12-12    ETTITGGTVARSVSTITRLFAPGASQN
101086-12-201086-12-    ETTITGGTAARSVSTFTRLFAPGASQK
1010-1418_M13uni_--_    ETTITGGTAARSVSTFTRLFAPGASQK
          101086-16-9   ETTITGGTAARSVSTFTKLFAPGASQK  5
        101086-12-14    ETTITGGTAARSVSTFTRLFAPGASQN
          101086-12-6   ETTITGGTAARGVRSFTRLFAPGASQN
101086-16-101086-16-    ETTISGGSAARSVYTFTGFFSPGAKQK
          101086-16-4   ETTISGGSAARSVYTFTGFFSPGAKQK
101086-16-1001086-16    ETTISGGSAARSVYTFTGFFSPGAKQK  3
M-16-13_M13uni_--_19    ETTISGGSAARSVYTFTGFFSPGAKQK
M-16-8_M13uni_--_9..    ETTISGGSAARSVYTFTGFFSPGAKQK
1010-1410_M13uni_--_    TTTITAGTAARGARSFVNLFSPGASQN
M-16-17_M13uni_--_7.    TTTITAGTAARGARSFANLFSPGASQN
1010-1416_M13uni_--_    ETTITGGTAAHSVRTIAEWFAPGASQK
M-16-9_M13uni_--_647    ETTISGGSAAHSVRTIADWFAPGASQK  2
101086-16-1101086-16    ATIITGGTAAHSVRTIADWFAPGASQK
M-16-1_M13uni_--_8..    ATIITGGTAAHSVRTIADWFAPGASQK  4
            Consensus   eTtIsGG.aArgv.sft.lF.pGA.Qn
```

6.3.4 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (9 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



6.3.4 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 10 -green, Week 12 – red, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Where two identical sequences occur the labels overlap.



6.3.4 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Time order phylogeny is clearly illustrated with sequential changes in the dominant HVR1 QS.

6.3.5 Results Summary

	Duration	HVR1 Substitution rate ($x10^{-5}$)	E1 Substitution rate ($x10^{-5}$)	Time Ordered Phylogeny	Number of Subpopulations	Cirrhosis	Change in dominant Subpopulation	Change in dominant during 16 weeks pre treatment	New subpopulation in Retrospective Sample
A	490	6.578	2.318	N	3	Y	N	N	Y
B	980	8.069	0.4695	N	4	Y	N	N	N
C	3715	3.435	2.393	N	1	Y	N	N	N
D	364	17.68	29.63	N	1	Y	N	N	N
F	275	7.777	0.9195	N	5	N	Y	N	N
G	234	5.245	2.015	Y	10	N	Y	Y	Y
H	288	9.846	5.681	Y	11	N	Y	Y	N
I	280	8.088	8.447	N	3	N	N	N	N
J	360	13.96	12.4	N	9	N	Y	Y	Y
K	412	18.51	8.01	Y	3	N	Y	Y	N
M	399	17.43	1.092	N	4	N	Y	N	Y
N	577	3.825	1.395	N	4	N	N	N	Y
O	287	19.67	8.544	Y	6	N	Y	Y	Y
Q	490	26.73	9.754	Y	9	N	Y	N	Y
R	761	3.677	6.937	N	3	N	Y	N	Y
U	399	23.86	5.255	N	2	N	N	N	N
V	574	8.169	4.52	N	5	N	Y	Y	N
W	444	4.192	3.651	N	2	N	N	N	N

6.3.5. Table 1. Summary of nucleotide substitution rates and subpopulation change. Duration indicates the timing pre treatment of the retrospective sample.

6.3.5.1 Time order Phylogeny

The inclusion of the retrospective sample increased the time period studied for the 18 subjects from 112 days to a median of 405 days (range 234-3715). When the additional sample was included the finding of a time order phylogeny was confirmed in all five subjects. In subject H however the dominant subpopulation in the pre treatment sample which had not been present in the sample 16 weeks pre treatment was also identified in the retrospective sample. Examination of the multiple sequence alignment of the amino acid HVR1 sequences for subject H however identifies two amino acid substitutions in the constituent sequence which has facilitated this re emergence (6.3.3 Fig. 6).

None of the four cirrhotic subjects demonstrated a change in the dominant subpopulation during the study. In contrast 10 of the 14 non cirrhotic subjects (71%) included demonstrated a change in the dominant subpopulation suggestive of a time order phylogeny. The difference between cirrhotic and non cirrhotic subjects was statistically significant (Chi^2 $p<0.05$) despite that fact that cirrhotic subjects were investigated over a longer period of time (median 735 days for cirrhotic subjects versus 399 for non cirrhotic subjects).

6.3.5.2 HVR stasis

Only two subjects were characterised by the presence of a single subpopulation for the entirety of the study including the retrospective sample, of which both were cirrhotic. Subject C was studied for in excess of 10 years and over this period of time the dominant HVR sequence differed from the original dominant sequence by a mere 2 amino acid substitutions. Subject D is characterised by a single dominant HVR1 QS subpopulation but with many different sequences within that subpopulation. Analysis of the phylogeny (Appendix A D.1 Fig 5 and Fig 6) indicates some change in the prevalence of individual sequences but the dominant sequence at the pre treatment sample was present in low number in the retrospective sample which was one year prior to commencing treatment.

Two non cirrhotic subjects (Appendix A U.1 Fig 7) and W (Appendix A.1 Fig 7)) were shown to be dominated in all samples by a single dominant HVR1 QS subpopulation but with the presence in minor copy of a second subpopulation on two occasions. These subjects are noteworthy for the fact that the dominant HVR1 sequence does not change by a single amino acid substitution over the 399 and 444 days they were studied respectively.

6.3.5.3 Number of Subpopulations

We evaluated the retrospective sample for the presence of a new subpopulation which had not been seen in the samples for the 16 weeks preceding treatment and found new subpopulations in 8 of the 18 subjects studied (6.3.5 Table 1). A new subpopulation was identified in a single cirrhotic patient (20%) and 50% 7/14 of non cirrhotic subjects had a new subpopulation present in the retrospective sample though this was not statistically significant ($p=0.34$ (χ^2)).

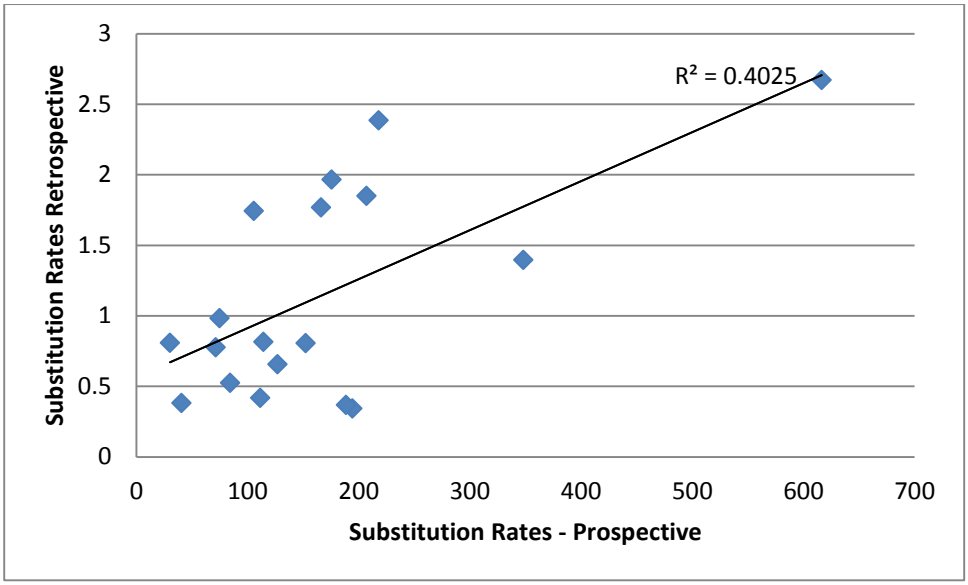
Although non cirrhotic patients had a higher mean number of subpopulations identified (5.42) during the study when compared with cirrhotic subjects (2.25), the difference was not statistically significant ($p=0.26$(χ^2)) though the limited number of participants may in part explain this.

6.3.5.4 Nucleotide Substitution Rates

	HVR1 Nucleotide Substiutions per site per year x10^{-4} Prospective Study	HVR1 Nucleotide Substiutions per site per year x10^{-4} including Restrospective
A	127	0.6578
B	152.2	0.8069
C	194.2	0.3435
D	166	1.768
F	71.22	0.7777
G	84.23	0.5245
H	74.7	0.9846
I	30.26	0.8088
J	348	1.396
K	206.9	1.851
M	105.6	1.743
N	40.49	0.3825
O	175.6	1.967
Q	616.4	2.673
R	188.6	0.3677
U	217.9	2.386
V	114.2	0.8169
W	111.3	0.4192

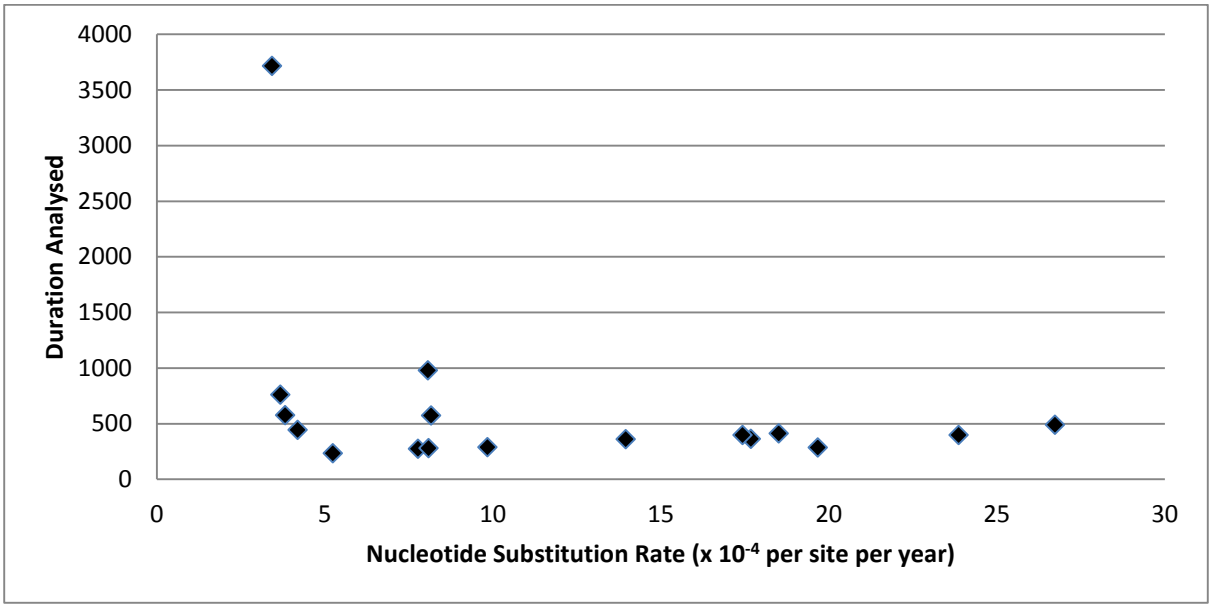
6.3.5.4 Table. 1 Comparison of HVR1 substitution rates between prospective study and when retrospective sample is included

The samples were interrogated using Beast© to calculate nucleotide substitution rate per site per year for HVR1. When compared with the figures calculated using the prospective samples over a 16 week period it became clear that there was significant disparity between the results obtained. The results for the 16 week study were in all cases several orders of magnitude greater than those when the retrospective sample was included.



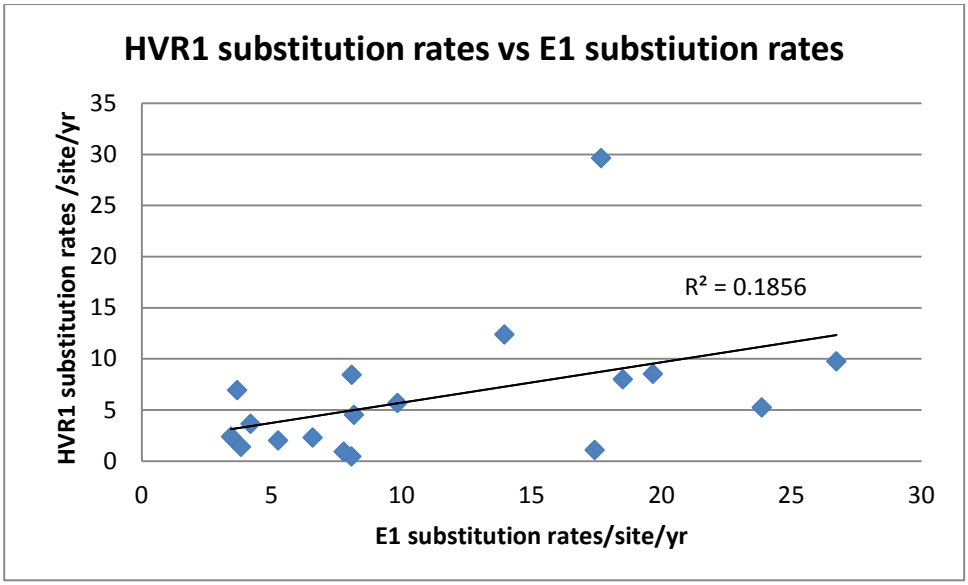
6.3.5.4 Fig. 1 Scatter plot of nucleotide substitution rates per year per site from prospective study and including retrospective sequences

We investigated for a relationship between the nucleotide substitution rates as calculated with and without the retrospective samples and found an association between these data (paired Student's t-test <0.01).



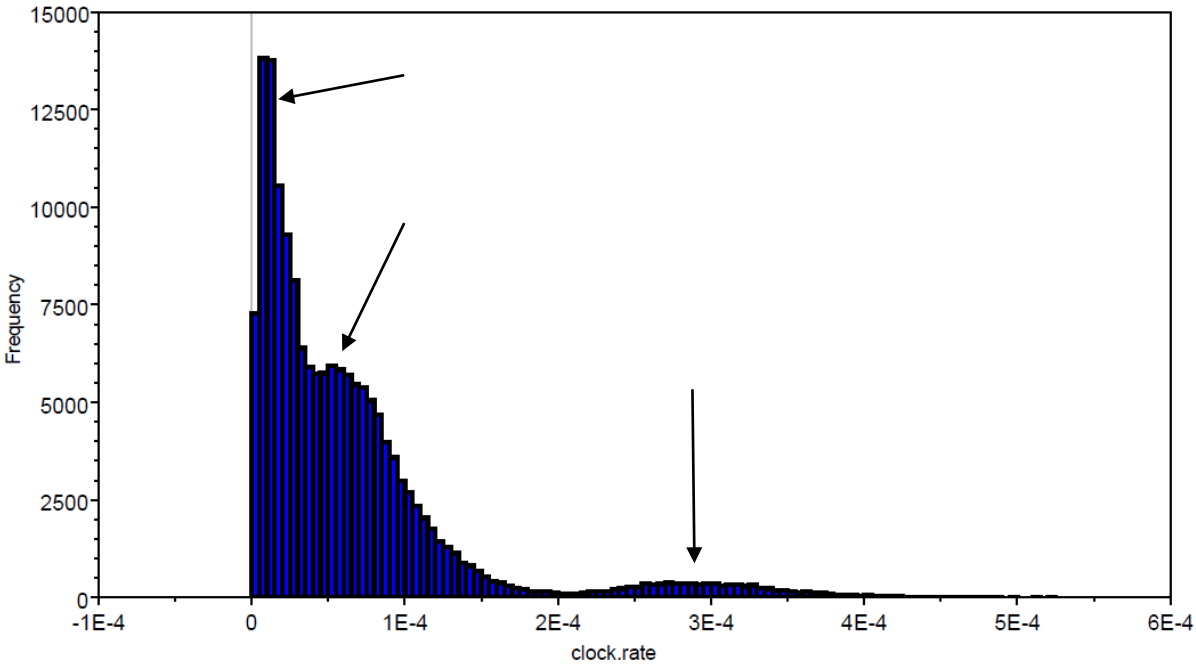
6.3.5.4 Fig. 2 Comparison of HVR1 substitution rates when duration of retrospective investigation is included.

There was no association between the length of time studies and the substitution rate.



6.3.5.4 Fig. 3 Comparison of HVR1 substitution rates with E1 substitution rates using retrospective clones.

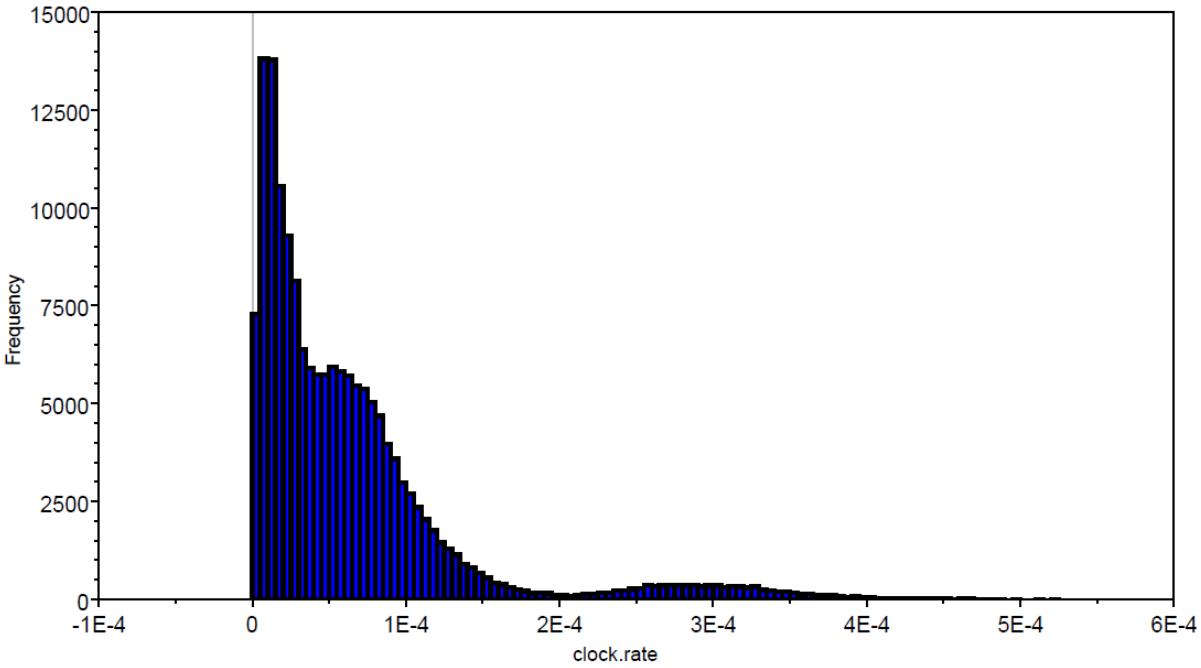
There was no correlation between HVR1 and E1 nucleotide substitution rates.



6.3.5.4 Fig. 4 Mean HVR1 substitution rates per nucleotide per site as calculated using Bayesian Evolutionary Analysis with 10,000,000 calculations for each subject.

We analysed the data produced from the substitution calculation by combining the result for all subjects in order to evaluate for patterns of substitution rate among all subjects. This produced a

skewed curve which suggested two or possibly three different average mutation rates as indicated with arrows.



6.3.5.4 Fig. 5 Mean E1 substitution rates per nucleotide per site as calculated using Bayesian Evolutionary Analysis with 10,000,000 calculations for each subject.

We further investigated the sequences to look for substitution rates in the E1 section and identified a similar curve. We compared the substitution rates for both E1 and HVR1 and there was evidence of correlation (paired two tailed Students T test $p<0.01$). This suggested that the underlying substitution rates for both E1 and HVR1 were related. We evaluated our E1 sequences for evidence of selection using REL (www.datamonkey.org) and found no evidence of either positive or purifying selection at individual codon sites in E1 for any of the subjects suggesting that all change was as a result of genetic drift.

E1 substitution rates did not correlate with cirrhosis, number of subpopulations, presence of multiple subpopulations, change in dominant subpopulation, and evidence of time ordered phylogeny.

6.4.1. Discussion

We have previously described in detail the change in HVR1 QS at two to four week intervals in a cohort of 23 subjects chronically infected with HCV. Our results highlighted a degree of HVR1 change which had not previously been described in the literature over such short time intervals(302). We identified clonal depth as a potential confounding factor in our previous analysis. Additionally, we postulated that natural variations in the prevalence of differing subpopulations rather than true selective change in sequences could in part explain the changes we described. Here we have used sequences generated from retrospective samples contained in our HCV library to confirm our previous findings.

6.4.2 Diversity, Complexity, and Divergence

Both HCV HVR1 diversity and complexity have previously been associated with likelihood of treatment success (341, 344, 346, 350, 354). We have previously shown that changes in these metrics of quasispecies are unpredictable and as a result likely to be of limited utility in predicting response to dual therapy with pegylated interferon and ribavirin(302). The inclusion of a retrospective sample has not altered this conclusion.

Diversity in E1 is limited in most cases and HVR1 diversity is only similar to E1 diversity in those subjects with low overall diversity. Diversity may provide some information with regard to the potential for the presence of multiple subpopulations is but is otherwise of limited utility in describing QS populations.

Complexity also provides limited information with regard to a QS population. It is possible by analysing the relative proportions of amino acid to nucleotide complexity at both E1 and HVR1 level to deduce some information with regard to whether the underlying QS may be under positive or purifying selection or following a path of genetic drift with little selective pressure. Nevertheless, the usefulness of this metric in describing the behaviour of QS limited by the cumbersome nature of how it is calculated and the ready availability of more useful tools for describing QS populations and change.

Divergence is perhaps the most useful of the three commonly used metrics of QS populations. The inclusion of retrospective samples has allowed us to confirm persistently low divergence among a number of predominantly cirrhotic subjects from the prospective arm of the study. E1 divergence was minimal in all subjects whereas HVR1 divergence was dramatic in many subjects. The magnitude of HVR1 divergence between the retrospective sample and the week 16 sample was often greater than that described in the fortnightly samples taken thereafter but the magnitude did not appear time dependent. Often the divergence between fortnightly samples approximated those seen between the retrospective sample and the week 16(pre treatment) sample. This suggests ongoing dramatic change

in the HVR1 QS. One possible explanation for this might be changes in the proportions of the differing subpopulations within the milieu(282).

When we looked at divergence from the retrospective sample this became somewhat clearer. As divergence was now being calculated from a static parameter, the noise generated by changes in the proportions of subpopulations could be silenced to a degree. In a number of subjects the magnitude of divergence calculated increased with each subsequent sample. This suggests truly divergent sequence change from the initial group of sequences. In some subjects, the initially significant divergence between the retrospective and week 16 sample was followed by a pattern of increasing and decreasing divergence from the retrospective sample. This has two possible explanations which require further analysis to confirm. Firstly, it may reflect differing proportions of a group of clones closely related to the retrospective sequences within the sample. Secondly, it may suggest multidirectional exploration of the sequence space within a multiple fitness optima. The virus in this circumstance is "testing" the various mutations available for mutants with significant fitness benefit in a process termed "pacing the cage"(408).

6.4.3. Phylogenetics

The visualisation of the HVR1 QS using phylogenetic trees clearly illustrates the degree of change from the retrospective sample and throughout the prospective portion of the study. The inclusion of scale bars provides additional valuable information with regard to the degree of HVR1 change observed. In Subject A (6.3.1 Fig. 5) the trees clearly suggest that the HVR1 QS has undergone significant change in the recent past which was not possible to appreciate using the prospective samples alone. Additionally, it identifies a likely ancestral divergence event which gave rise to the group of clones seen at week 14 (grey circles in the bottom right clade).

The phylogenetic tree for subject C (6.3.1. Fig. 5) illustrates sequence stasis and a complex and diverse but closely related QS pattern.

These patterns contrast markedly with the tree produced for subject Q (6.3.4. Fig. 3) where the scale bar suggests a markedly different magnitude of change. Nonetheless, even with scale bars it can be difficult to appreciate the degree of change illustrated in a phylogenetic tree.

6.4.4 Sequence Alignments and Subpopulations

Sequence alignments are an unrefined way of presenting QS data but we include raw amino acid sequence alignments of the HVR1 for a number of reasons. Firstly, phylogenetic trees cannot give an impression of the underlying sequence change. Secondly, they may facilitate the division of the QS into groups of sequences that are more closely related to each other called subpopulations.

We have previously defined subpopulations as groups of sequences that differ from each other by less than four amino acid substitutions (15% of HVR1) and from all others within the sample by four or more amino acid substitutions(302). Subpopulations inform with regard to the extent of QS change and are useful in identifying subjects with large scale change in the QS. The use of prevalence graphs further illustrates temporal changes in subpopulations.

Here for the first time we have presented how superimposing the subpopulations identified using sequence alignments on the phylogenetic trees validates this method for describing evolutionary change. Using this strategy, we had described a time order phylogeny in five of the subjects and the inclusion of retrospective sequences confirms this finding. Therefore, we are satisfied our use of clonal sequences accurately identifies time order phylogeny over periods as short as 16 weeks. 60% of the subjects where time order phylogeny was confirmed had a novel subpopulation in the retrospective sample (6.3.5. Table 1). This suggests that the rapid change identified in our cohort may be sustained over prolonged periods of time and these differing patterns of QS change have not previously been defined in the literature.

Novel subpopulations were described in the retrospective sample of 70% of subjects and that this resulted in a novel time order phylogeny in four additional subjects (A, F, J, M). Conversely, the inclusion of retrospective samples confirmed QS stasis in four subjects (C, D, U, W). Subject C is perhaps the most interesting of these as we were in a position to interrogate samples covering a period of 10 years prior to entry into the prospective study. This patient had been infected with HCV through the use of contaminated Anti D immunoglobulin in the post partum management of women who were Rhesus antigen negative in 1977 in Ireland(41). In the 10 years prior to study entry, our analysis has demonstrated minimal change in the HVR1 QS suggesting a virus well adapted to the host immune system.

6.4.5 Substitution Rates

The inclusion of retrospective samples allowed us to investigate the estimated substitution rate per nucleotide per year over more prolonged period of time (median 405 days). Our data for HVR1 substitution rates was on average 10 to 100 times greater than the reported mutation rate for HCV. In noting this we acknowledge that HVR1's role as an immune target would mean that the substitution rate would appear far greater than the underlying mutation rate. The mean substitution rate for HVR1 from the literature is of the order of 7×10^{-3} per nucleotide site per year which contrasts with our mean of 17×10^{-3} per nucleotide site per year from our prospective data(407). The substitution rates calculated in this paper were however calculated using 31 whole genome sequences derived from 15 different subjects and do not give any information with regard to the underlying QS. Our analysis

provides for the first time estimates of substitution rates based on groups of clones from different time points. Nevertheless we were eager to explore whether such high substitution rates would be found when the same individuals were investigated over longer time intervals.

The inclusion of retrospective samples reduced the mean substitution rate per site per year significantly from 17×10^{-3} to 11.5×10^{-5} for HVR1. We had not anticipated such a dramatic fall but on review of the data we feel that the initial calculation for the prospective 16 week study overestimated the substitution rate. This we feel was because the ancestral sequences used in the calculation were groups of sequences with variable diversity. The software then calculated substitution rates on the basis that sequences were all progeny of the original group of sequences. In order to confirm this we have subsequently investigated for an association between the number of subpopulations described for each individual and the HVR1 substitution rate as calculated using BEAST. Using a paired two tailed Student's t-test we can demonstrate an association between the number of subpopulations described and the substitution rate calculated ($p<0.0001$).

This highlights a potential pitfall in calculating substitution rates which is difficult to overcome – particularly in the setting of subjects with multiple subpopulations that are sustained over prolonged periods of time.

We examined the substitution rates calculated using the retrospective samples for similar correlation and found a weaker association (Pearson 0.137, paired two tailed Student t-test $p<0.02$). This suggests that the time interval reduces the effect of the original set of clones on the overall substitution rate. Furthermore, we investigated for an association between the length of time the virus was studied and found a weaker correlation with the substitution rate as calculated using BEAST ($p<0.05$).

Combining these findings leads us to conclude that the HVR1 substitution rate that is most likely to reflect the overall mutation rate would be identified in subjects with the fewest subpopulations and studied over the longest time interval. Subject C is characterised by a single subpopulation and is studied over 10 years and has an E1 and HVR1 substitution rate of 2.393 and 3.435×10^{-5} per site per year respectively. E1 in this subject is under no identifiable selective pressures implying that this figure may be a true reflection of the underlying HCV substitution rate.

6.4.6 E1 vs HVR1 Substitution Rates

HVR1 is well recognized as an immune target with a malleable structure including many potential epitope binding sites and is thought to act as a decoy protecting more structurally constrained portions of the HCV envelope protein(310, 313). E1 is not a recognised immune target and accordingly

we found no correlation between HVR1 substitution rates and those calculated using E1 alone (6.3.5.4 Fig. 3).

Graphical representation of estimated substitution rates as calculated using BEAST for all subjects suggest three likely substitution rates (6.3.5.4 Fig. 4). Our data may suggest that the virus has differing underlying mutation rates in different subjects. The duration of infection is unknown in many cases but certainly subject C had a prolonged duration of chronic infection and was also demonstrated to have one of the lowest substitution rates. We have previously postulated that the underlying mutation rate of a polymerase may be a selectable trait depending on the requirement placed on the virus to adapt to host environment(409). Although our numbers are limited we feel that the similar pattern of substitution rate seen in E1 compared to HVR1 supports this hypothesis.

6.5. Conclusion

The inclusion of retrospective samples has confirmed the patterns of HVR1 QS change seen in the prospective study. We found no evidence that the time order phylogeny was incorrectly identified in the five patients with this pattern of change who were included. We found evidence for time order phylogeny in a further four subjects when the period under review was extended. QS diversity, complexity, and to a lesser degree divergence are of limited use in describing QS though a systematic approach to their interpretation may provide some insights. QS subpopulations are a useful tool for identifying and describing widespread change in QS. Despite high substitution rates, HCV is able to maintain single subpopulation infection with minimal amino acid substitutions over decades. The calculation of underlying HVR substitution rates remains challenging though the inclusion of estimates of E1 substitution rates can be useful in deciphering the underlying mutation rate in chronic HCV infection. HCV may have a variable mutation rate which decrease with time as the virus adapts to the host and the requirement for immune escape diminishes once niche deficits in the humoral immune system have been exploited.

Chapter 7

Published Study

Network analysis of the chronic Hepatitis c virome defines HVR1 evolutionary phenotypes in the context of humoral immune responses

Chapter contributors:

Proposal

Next generation sequencing project planned by Daniel Schmidt-Martin with samples chosen based on analysis performed on cloning study described in Chapter 4.

Laboratory

Samples collected, stored and nested PCR product generated, purified and prepared for 454 sequencing by Daniel Schmidt-Martin.

All data from prospective cloning study generated by Daniel Schmidt-Martin.

Imunoglobulin fractionation of stored samples performed by Brendan Palmer

Bioinformatics

Quality control, and analysis of next generation sequencing outputs performed by Brendan Palmer.

Paper drafting

Paper written and edited by Brendan Palmer and Daniel Schmidt-Martin with input from all contributing authors under the supervision of Liam Fanning.

Figures created by Brendan Palmer.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Network Analysis of the Chronic Hepatitis C Virome Defines Hypervariable Region 1 Evolutionary Phenotypes in the Context of Humoral Immune Responses

Brendan A. Palmer,[a] Daniel Schmidt-Martin,[a] Zoya Dimitrova,[b] Pavel Skums,[b] Orla Crosbie,[c] Elizabeth Kenny-Walsh,[c] Liam J. Fanning[a]

Molecular Virology Diagnostic & Research Laboratory, Department of Medicine, University College Cork, Cork, Ireland[a]; Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[b]; Department of Hepatology, Cork University Hospital, Cork, Ireland[c]

ABSTRACT

Hypervariable region 1 (HVR1) of hepatitis C virus (HCV) comprises the first 27 N-terminal amino acid residues of E2. It is classically seen as the most heterogeneous region of the HCV genome. In this study, we assessed HVR1 evolution by using ultradeep pyrosequencing for a cohort of treatment-naive, chronically infected patients over a short, 16-week period. Organization of the sequence set into connected components that represented single nucleotide substitution events revealed a network dominated by highly connected, centrally positioned master sequences. HVR1 phenotypes were observed to be under strong purifying (stationary) and strong positive (antigenic drift) selection pressures, which were coincident with advancing patient age and cirrhosis of the liver. It followed that stationary viromes were dominated by a single HVR1 variant surrounded by minor variants comprised from conservative single amino acid substitution events. We present evidence to suggest that neutralization antibody efficacy was diminished for stationary-virome HVR1 variants. Our results identify the HVR1 network structure during chronic infection as the preferential dominance of a single variant within a narrow sequence space.

IMPORTANCE

HCV infection is often asymptomatic, and chronic infection is generally well established in advance of initial diagnosis and subsequent treatment. HVR1 can undergo rapid sequence evolution during acute infection, and the variant pool is typically seen to diverge away from ancestral sequences as infection progresses from the acute to the chronic phase. In this report, we describe HVR1 viromes in chronically infected patients that are defined by a dominant epitope located centrally within a narrow variant pool. Our findings suggest that weakened humoral immune activity, as a consequence of persistent chronic infection, allows for the acquisition and maintenance of host-specific adaptive mutations at HVR1 that reflect virus fitness.

Hepatitis C virus (HCV) infection is a global health issue and is recognized as a major etiological agent of liver-related diseases (1). It has been estimated that the current prevalence of HCV represents approximately 2% of the global adult (15 years of age and older) population (2). Following transmission, HCV infection may remain asymptomatic for decades, resulting in the majority of infections initially passing undetected (3). It is estimated that up to 4 million Americans are living with the virus, the majority of whom became infected prior to the isolation and identification of the virus (4, 5). Consequently, the U.S. Centers for Disease Control and Prevention now recommend that Americans born from 1945 to 1965 be screened for the presence of the virus notwithstanding the presence of clinical symptoms (3, 5).

HCV is a single-stranded positive-sense RNA virus of considerable genomic heterogeneity. A recent reclassification defined the HCV global distribution into 7 genotypes and 67 subtypes, with genotypes 1 and 3 accounting for the majority of infections worldwide (6, 7). An error-prone RNA-dependent RNA polymerase, together with an inherent tolerance of defined hypervariable regions (HVR), accounts for much of this variability. Three HVRs are located within the envelope glycoprotein E2. The greatest heterogeneity has been identified at the 27-amino-acid HVR1 (residues 384 to 410 of the H77 reference strain), located at the amino-terminal end of the E2 glycoprotein (8). Recent studies indicated that the central region of E2 (residues 456 to 656) is globular and surprisingly compact, whereas the first 80 amino acids (including HVR1) lack this structural rigidity (9). This observation is consistent with a region that is proposed to shield conserved neutralizing epitopes and to participate in high-density lipoprotein enhancement of infection via scavenger receptor class B type I (SRBI) interactions and is itself targeted by neutralizing antibodies (nAb) (10–16).

Mutational flexibility at HVR1 was characterized soon after the initial identification of HCV (8, 17). Rapid mutational change of HVR1 has been documented over weeks during the acute phase of infection, where HVR1 evolution is governed predominantly by strong selective pressures, with fixation of beneficial mutations (11, 18, 19). Reports examining samples collected over years to decades have documented the emergence of convergent HVR1

Received 25 November 2015 **Accepted** 22 December 2015

Accepted manuscript posted online 30 December 2015

Citation Palmer BA, Schmidt-Martin D, Dimitrova Z, Skums P, Crosbie O, Kenny-Walsh E, Fanning LJ. 2016. Network analysis of the chronic hepatitis C virome defines hypervariable region 1 evolutionary phenotypes in the context of humoral immune responses. J Virol 90:3318–3329. doi:10.1128/JVI.02995-15.

Editor: M. S. Diamond

Address correspondence to Liam J. Fanning, l.fanning@ucc.ie.

B.A.P. and D.S.-M. contributed equally to this article.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

TABLE 1 Study cohort descriptors[a]

Patient	Virus genotype	Age (yr)	Sex	Baseline viral load ($\log_{10}$ IU/ml)	Mode of transmission	Cirrhosis	Group identifier
1	1a	47	M	5.37	Unknown	Y	ML2
2	1a	43	F	6.40	Unknown	N	ML3
3	1b	75	M	6.08	Blood transfusion	Y	ST1
4	1b	50	M	6.91	Intravenous drug use	Y	ST2
5	1b	59	F	6.59	Blood product	Y	ST3
6	1b	61	M	6.61	Unknown	Y	ST4
7	3a	41	M	6.57	Iatrogenic	N	ML1
8	3a	37	F	4.70	Intravenous drug use	N	AD1
9	3a	21	M	4.91	Iatrogenic	N	AD2
10	3a	23	F	6.28	Unknown	N	AD3
11	3a	45	M	5.18	Unknown	Y	ST5
12	3a	32	M	6.74	Unknown	N	ST6

[a] F, female; M, male; Y, cirrhosis was present; N, cirrhosis was not present.

quasispecies variant pools under purifying selection pressures in established chronic infections (20–24). In selected instances, the maintenance of the dominant HVR1 epitope extended over years and in the absence of an associated antibody response (22).

We recently reported HVR1 quasispecies phenotypes at the clonal level from a study of 23 chronically infected, treatment-naive patients from whom samples were collected every 2 weeks over a period of 16 weeks (25). Within the short sampling time frame, both stationary (ST) viromes and rapid intrapatient sequence changes were observed. In the present study, a representative cohort of 12/23 patients was selected for ultradeep pyrosequencing (UDPS) analysis to interrogate in depth the clonal phenotypes reported. Furthermore, IgG-associated virions were subfractionated from serum, and the HVR1 profiles of viral RNA-positive samples were determined. We report HVR1 phenotypes exhibiting conservative HVR1 evolution that is coincident with patient age and the presence of cirrhosis. The HVR1 variant pools of this group were interlinked by single-site amino acid substitutions. Additionally, IgG binding for this cohort of patients was associated with the dominant HVR1 variant but was not indicative of effective virus neutralization for the majority of patient viromes during the study period.

MATERIALS AND METHODS

Sample set. Twelve treatment-naive patients were selected from a larger cohort of 23 for whom the HVR1 quasispecies change had been characterized at the clonal level (Table 1) (25). The selection criteria were based on genotype, the presence/absence of cirrhosis, and the observed divergence of sequences from the initial sample observed from clonal analysis (25).

Samples collected at 16 weeks, at 8 weeks, and immediately prior to the commencement of antiviral treatment (0 weeks) were analyzed in conjunction with a homogenous plasmid control template of known sequence (GenBank accession number GQ985374) by UDPS of amplicons that spanned the E1-E2 glycoprotein gene junction. The amplified fragment corresponded to positions 1296 to 1613 of the H77 reference strain (GenBank accession number AF009606). The 0-week sample was additionally analyzed for the presence of IgG-bound virions. Subjects provided written informed consent, and the study was undertaken under the governance of the Clinical Research Ethics Committee of the Cork Teaching Hospitals.

Fractionation of IgG-bound virions. Protein G HP SpinTrap columns were used to extract IgG-bound virions from whole patient serum (GE Healthcare). The procedure followed the manufacturer's instructions, with minor modifications. Briefly, 200 μl of patient serum, normalized to 5 $\log_{10}$ IU/ml with phosphate-buffered saline, was applied to a preequilibrated SpinTrap column. End-over-end mixing at room temperature (RT) for 15 min was then performed, followed by centrifugation. Eight serial wash steps (W1 to W8) with 300 μl of binding buffer were applied. The final wash elute (W8) was retained and analyzed to confirm the absence of detectable virus RNA. IgG-bound virions were recovered from the column by the addition of 200 μl of elution buffer followed by end-over-end mixing at RT for 5 min. The eluted IgG-bound virions were collected in a microcentrifuge tube containing 30 μl neutralizing buffer.

Amplification of the E1-E2 junction encompassing the HVR1 region. Viral RNA was extracted using a QIAamp viral RNA minikit into a final volume of 60 μl (Qiagen). Ten microliters of the RNA sample was taken to generate cDNA by use of SuperScript II reverse transcriptase (Invitrogen). Nested PCR amplification was performed as described previously (26). Inter- and intrapatient samples were handled on separate days to guard against cross-contamination. In each instance, a 1:100 dilution of the RNA was performed to ensure that the amount of starting template was not limiting. This was confirmed by visualization of the amplicon by gel electrophoresis. Amplicons were purified using a PCR purification kit (Qiagen).

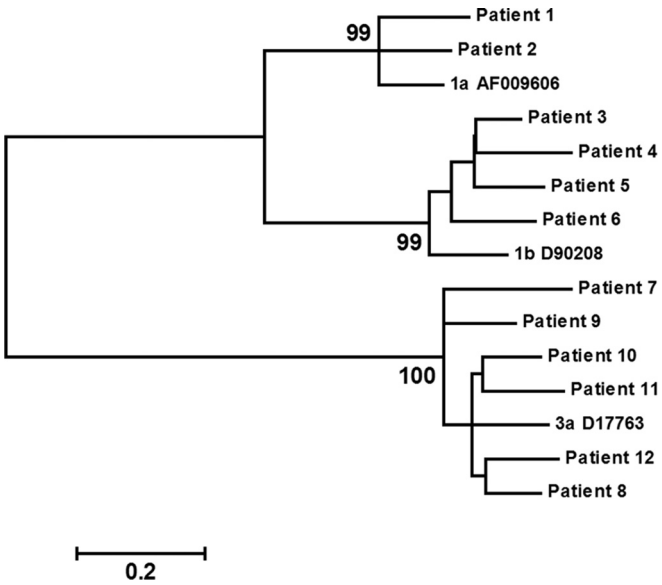


FIG 1 Phylogenetic analysis of patient consensus sequences against reference 1a, 1b, and 3a strains. The scale bar shows the genetic distance. Bootstrap values for 1,000 resamplings are shown.

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

TABLE 2 Inter- and intracomponent HVR1 epitope distribution

Group identifier	Group	Sublineage	Total lineage frequency[a] (%)	No. of 1-step components[b]	No. of unique HVR1 epitopes per 1-step component	No. of HVR1 epitopes jointly isolated through UDPS and clonal techniques[c]	UDPS HVR1 sample space captured by clonal analysis (%)
ST1	ST		100	1	10	2	98.15
ST2	ST		100	5	20, 1, 1, 1, 1	3	96.12
ST3	ST		100	1	28	3	95.33
ST4	ST		100	1	22	8	86.89
ST5	ST		100	1	2	2	100
ST6	ST		100	3	16, 1, 1	6	99.01
AD1	AD		100	4	8, 8, 1, 1	6	95.04
AD2	AD		100	9	6, 2, 2, 1, 1, 1, 1, 1, 1	6	98.88
AD3	AD		100	7	4, 2, 2, 1, 1, 1, 1	6	97.20
ML1	AD	L1	99.6	8	8, 7, 4, 3, 3, 1, 1, 1	11	91.13
	ND[d]	L2	0.4	1	2	2	0.34
ML2	ST	L1	88.4	4	19, 3, 2, 1	8	85.57
	ND	L2	11.6	2	3, 1	1	12.13
ML3	ST	L1	96.1	3	6, 1, 1	7	95.46
	ND	L2	3.9	1	1	1	3.88

[a] Averaged across all three samples.
[b] All HVR1 variants within any one component can be linked to at least one other variant within that component by a single amino acid substitution.
[c] See reference 25.
[d] ND, not determined.

Clonal analysis. IgG-bound virion RNA was isolated and the E1-E2 region amplified. Amplicon-positive samples were initially purified and sequenced (Eurofins Genomics). In cases where multiple peaks were observed in the trace files, a panel of clones was generated as previously described (26).

UDPS data generation, handling, and error correction. Amplicons were quantified using a Biophotometer machine (Eppendorf) and diluted to a final concentration of 1×10^7 molecules/ml. Pyrosequencing was performed using a 454 GS FLX titanium platform with sample-specific multiplex identifier sequence-adapted libraries for Lib-1 sequencing (Roche 454 Life Sciences, Branford, CT). Raw sff data files were first uncoupled into individual patient sample files by using SFFFile tools (Roche). Low-quality reads and reads shorter than 90% of the expected amplicon lengths were removed.

The resultant data files were sequentially processed through implementation of the *k*-mer error correction (KEC) and empirical threshold algorithms as previously described, using the parameters $k = 25$ and $i = 3$ (22, 27). A panel of clonal sequences temporally matched to the UDPS data was used to further identify and correct homopolymer errors (22, 25). Following this procedure, no erroneous sequences were present at a frequency of >0.1% in the homogeneous plasmid control sample. Consequently, all haplotypes present at a frequency of >0.1% in their respective samples were retained for downstream analysis.

1-step and *k*-step network construction. To study the dynamics of intrahost quasispecies evolution, we created two networks for each patient (28). First, all unique haplotypes (318 bp) were aligned, and the Hamming distance between each pair was calculated. Connected components were then built, in which each unique haplotype was represented by a node and two nodes were connected by an edge if the distance between them was 1. Initially, the components were independent of one another and together formed a 1-step network.

The 1-step network of most patients consisted of several components. To join them together, *k*-step networks were constructed as follows: iteratively for $k = 2, 3, \ldots$, until all pairs of haplotypes from different components with a distance equal to *k* were found. They were linked by edges, and the components were recalculated. These steps were repeated until a single connected component was formed. The resulting *k*-step network is equivalent to the union of all minimum spanning trees. The analysis and network visualization were performed with MATLAB R2014b (The MathWorks, Inc.) and Pajek (29).

Bioinformatic analyses. MEGA6 was used to calculate Hamming distances and synonymous and nonsynonymous substitution rates (30). Phylogenetic trees were drawn in MEGA6, using a general time-reversible model with gamma-distributed and invariant sites. The time-ordered Shannon diversity (H) of 1-step networks was calculated using the following formula: $H = -\sum_{i=1}^{N} p_i \times \ln p_i$, where p_i is the total frequency of haplotype component i in the 1-step network and N is the number of components of the 1-step network (richness). The evenness (E_H) of the 1-step network was determined using the following formula: $E_H = H/\ln N$.

Three patients were identified as containing mixed lineages. In each instance, the components comprising the dominant lineage were analyzed separately from the minor lineages. Prior to calculation of the Shannon diversity index, the total frequency of the dominant lineage components was normalized to 1 to account for the absence of the minor lineage.

Amino acid conservation plots were drawn using the Jalview program, which is based on analysis of multiply aligned sequences (AMAS) to determine changes to the physiochemical properties of the constituent amino acids (31, 32).

Nucleotide sequence accession numbers. UDPS data sets used in this study are available at http://www.ucc.ie/liamfanning/hcv. Unique nucleotide sequences were deposited in GenBank and assigned accession numbers KT193821 to KT193838.

Statistical analysis. All statistical analyses were performed using R 3.1.3. The statistical significance of comparisons was analyzed using the nonparametric Mann-Whitney U test. Where appropriate, the nonindependence of intrapatient samples was accounted for by averaging the individual values, which were then used for statistical comparisons. In all tests, P values of <0.05 were considered statistically significant.

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

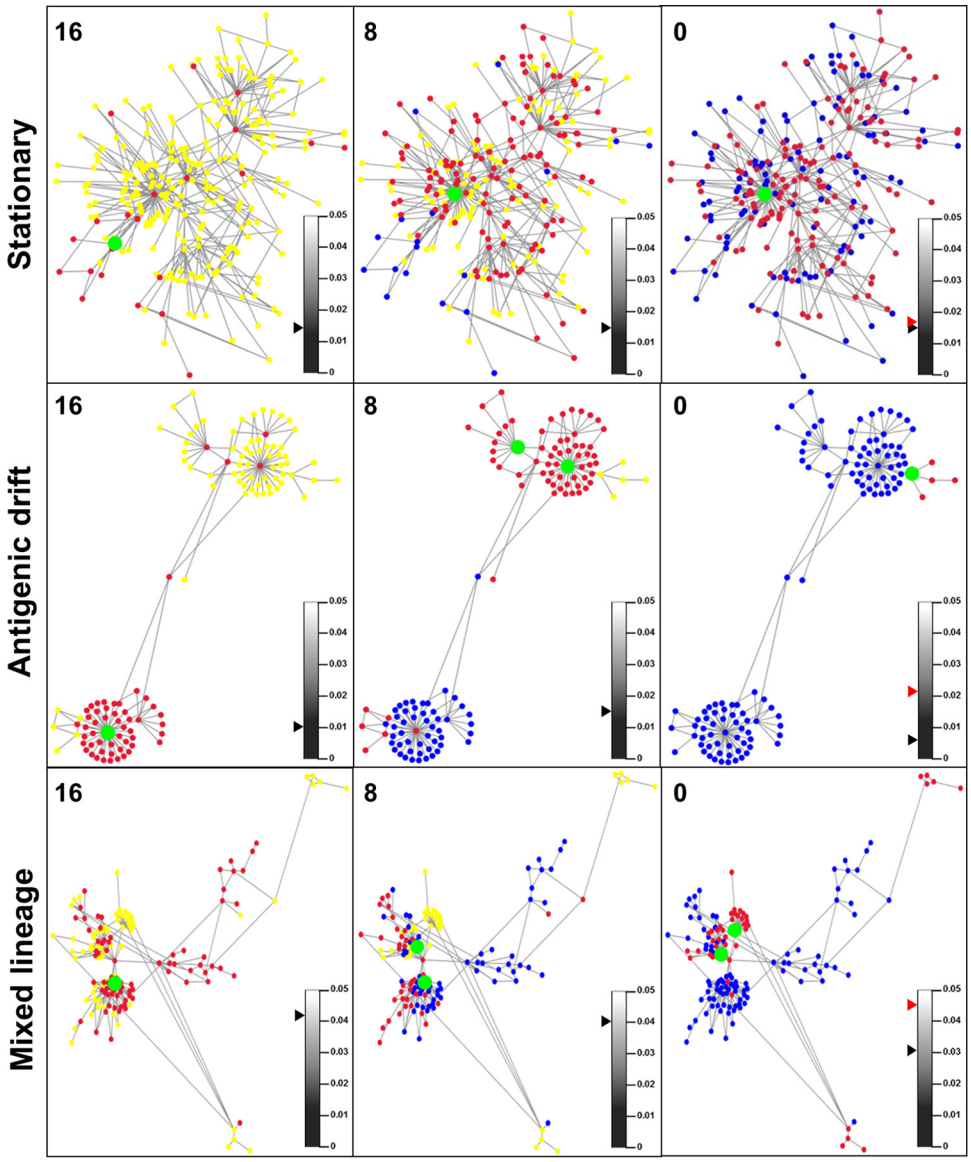


FIG 2 Representative *k*-step network graphs for stationary (ST3), antigenic drift (AD1), and mixed-lineage (ML1) viromes. Unique haplotypes with frequencies of >0.1% of the sample space are displayed. Individual panels are given for 16 weeks pretreatment, 8 weeks pretreatment, and immediately prior to commencement of treatment (0 weeks). Red nodes identify unique time point-specific haplotypes, yellow nodes identify future haplotypes, and blue nodes identify haplotypes from an earlier sample that were not detected in the given sample. Green nodes denote sample-specific haplotypes that occupied >10% of the sample space at that time. The maximum distance between any two haplotypes from separate 1-step components was 4 bp for ST3, 8 bp for AD1, and 18 bp for ML1. A black triangle juxtaposed to the color bar represents the relative Hamming distance for that samples. A red triangle represents the relative Hamming distance of haplotypes combined across all three samples.

RESULTS

Characterization of the patient cohort. Twelve treatment-naive patients chronically infected with either HCV genotype 1 ($n = 6$) or genotype 3a ($n = 6$) comprised the study cohort. Initial genotype identification was performed using the Versant hepatitis C virus genotype assay (LiPA) 1.0, targeting the 5′ untranslated region. This procedure identified all six genotype 1 patients as having subtype 1b virus (25). Reanalysis of both the 5′ untranslated region and the core by using LiPA 2.0 categorized patients 1 and 2 as having subtype 1a viruses. This result was confirmed by phylogenetic analysis of patient consensus sequences against reference 1a, 1b, and 3a sequences (Fig. 1 and Table 1).

In our hands, clonal analysis accounted for (on average) 96.3% occupation of the HVR1 variant sample space identified through UDPS (range, 86.9% to 100%). However, this translated to <34% of the unique HVR1 variant sequence space present in the UDPS data set (Table 2) (25). Frequency selection bias toward the dominant epitope in the clonal data was evident, as 34/35 HVR1 variants, with a UDPS sample-specific frequency of >5%, were also described clonally (25).

***k*-step network analysis of patient viromes.** To better explore sequence evolution over the sampling time frame, visualization of UDPS data was performed by generating a *k*-step network for each of the 12 patients. The initial 1-step network was comprised of

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

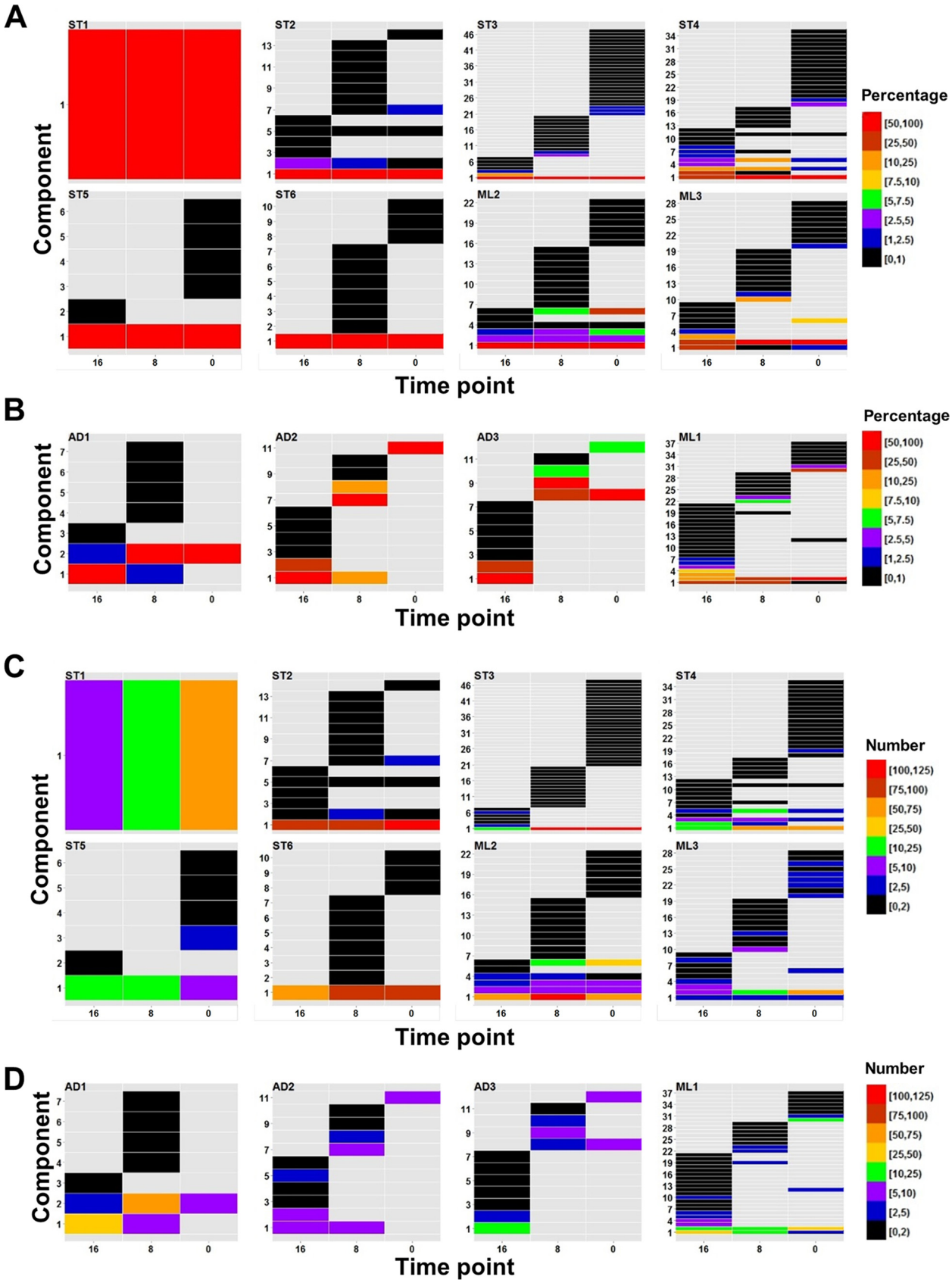


FIG 3 Temporal 1-step component frequency and composition. Over the 16-week sampling period, all components that formed the 1-step network graph were examined for fluctuations in the percentage of occupation of the sample space (stationary viromes [A] and antigenic drift viromes [B]) and the number of unique

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

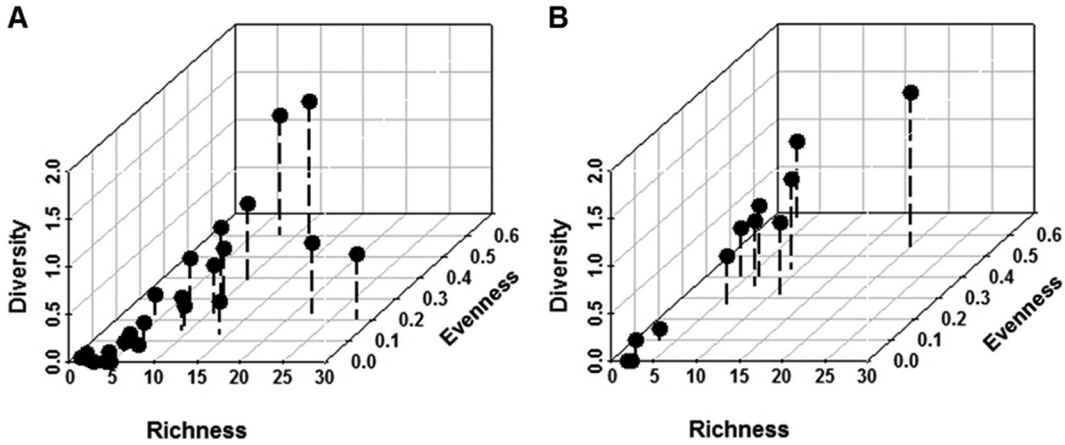


FIG 4 Richness, Shannon diversity, and evenness values for individual-sample 1-step networks. For each 1-step network ($n = 36$), the constituent component profile was assessed. (A) A total of 19/24 1-step ST networks were concentrated in the lower left quadrants for all three parameters examined. (B) A total of 6/12 1-step AD networks were concentrated in the lower left quadrants for all three parameters examined.

components whose nodes (haplotypes) were connected by edges that represented a genetic distance of 1. In cases where one haplotype could not be paired with a second in this manner, the component was comprised of that single haplotype.

The patient cohort was divided into three groups based on genetic and network characteristics, and representative *k*-step networks for each group are given in Fig. 2. First, viromes were classified as stationary (ST) based on comparable sample-specific and combined Hamming distances. Within ST networks, haplotype emergence over the sampling period was within a localized sequence space. The dominant haplotypes in each ST network remained largely fixed across the 16 weeks (Fig. 2, green nodes). Second, viromes were observed that exhibited a time-ordered spatial distribution of haplotypes toward naive sequence space. Much of the sequence heterogeneity was within HVR1, and such viromes were classified as undergoing antigenic drift (AD). The elevated Hamming distance of combined AD samples compared to sample-specific Hamming distances also defined the intersample heterogeneity. The remaining viromes contained mixed-lineage (ML) virus subpopulations (Tables 1 and 2). With respect to the ML phenotype, the presence of a subpopulation was initially evident due to elevated between-component distances (>16 bp) within the *k*-step network. These separations were later confirmed through a phylogenetic analysis using the maximum likelihood method, based on a general time-reversible model. Bootstrap values of >98% for 1,000 resamplings were recorded (data not shown). The dominant lineage in each of the ML viromes was partitioned away from minor lineage haplotypes and designated either ST or AD for downstream analysis. ML1 exhibited an antigenic drift phenotype. ML2 and ML3 were classified as stationary (Table 1). Minor lineages within each ML virome either were not detectable in all three samples or had too few unique haplotypes to be classified formerly as ST or AD (Table 2).

All 12 patients' *k*-step networks contained 1 to 4 dominant haplotypes (present at >10% of the sample sequence space) (Fig. 2, green nodes). Thirty-four of 36 samples contained a single haplotype that accounted for >25% of the sample space (the maximum recorded was 98.7%, for the AD1, 0-week sample). The dominant haplotype was centrally placed within the network and contained the highest edge degree (Fig. 2). The majority of edges to the dominant haplotype were reflective of synonymous substitution events.

1-step components reveal virome connectedness. For each patient network, the composition of the constituent components was analyzed over time (Fig. 3). ST patients were largely defined by a single dominant component that persisted across the 16-week sampling period and varied substantively only by the constituent numbers of unique haplotypes. The sequence depth achieved here facilitated the construction of a 1-step network (defined as being comprised of a single connected component) for ST1 sequences (Fig. 3). The majority of ST networks were formed from a single dominant component and multiple low-frequency components containing few unique haplotypes. In contrast, AD networks exhibited a temporal component dominance that was observed to change between sample points (Fig. 3B and D).

Partitioning of the patient data into 1-step components allowed for (i) haplotypes to be grouped together by the nearest evolutionary linkages and (ii) component stability over time to be assessed quantitatively. In this study, virome richness was defined as the number of components that accounted for all sequences within a sample-specific 1-step network. Shannon diversity and evenness are commonly used to characterize species diversity in a community. We applied this approach to our data by assuming that each component is analogous to a species and that, together, the components comprise the community. Evenness values approaching zero indicate a skewed component dominance within the network (e.g., ST1 is uneven because it contains only one component) (Fig. 3A). Overall, ST samples demonstrated restrictive exploration of the sequence space, which is indicative of component stability and dominance (Fig. 4). Nineteen of the 24 patient

haplotypes (stationary viromes [C] and antigenic drift viromes[D]). Each component was comprised of a single haplotype (black) or existed as a connected component where the constituent haplotypes differed by a single nucleotide substitution from at least one other haplotype across the length of the amplicon sequence (318 bp). Time is represented on the *x* axis, and the specific component identifier by patient is shown on the *y* axis. Ranges are defined by the color bars.

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

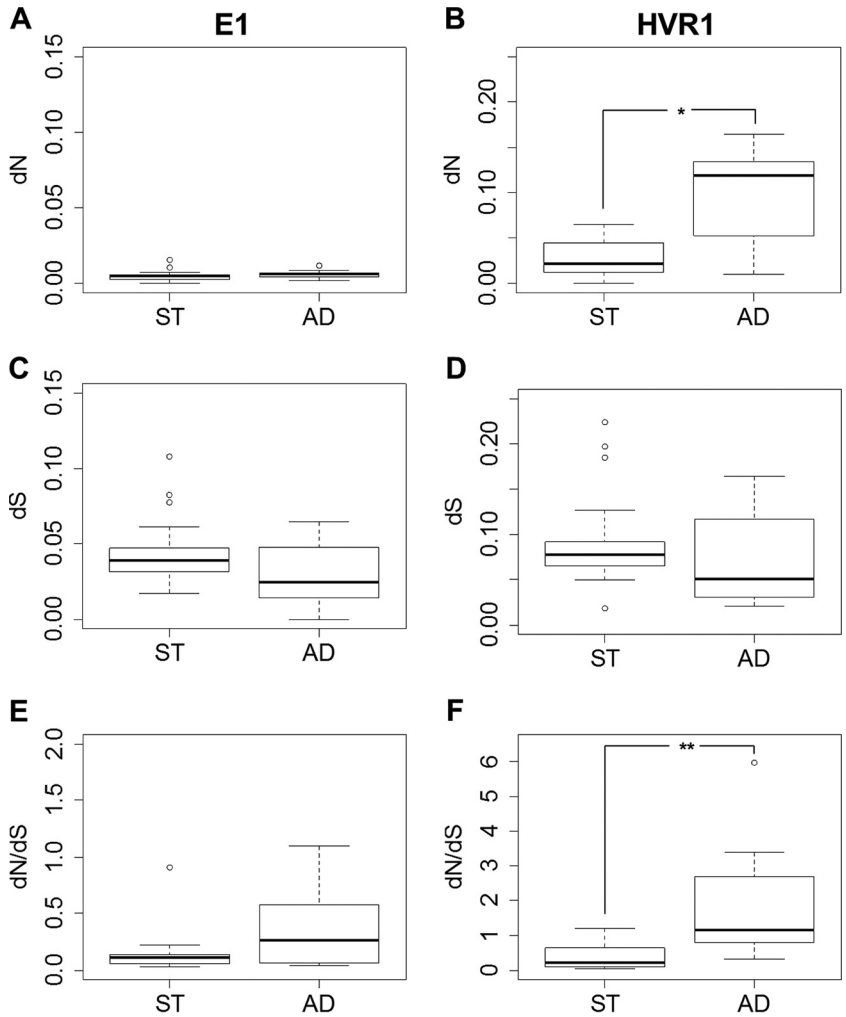


FIG 5 *dN*, *dS*, and *dN*/*dS* variations between patient groups ST and AD examined across E1 (195 bp) and HVR1 (81 bp) sections of the amplicon sequence. (A and B) Number of nonsynonymous mutations per nonsynonymous site. (C and D) Number of synonymous mutations per synonymous site. (E and F) *dN*/*dS* ratios. When a patient was identified as containing a mixed-lineage virome, the dominant lineage was assigned to either the ST or AD group, where possible (Table 2), and analyzed independently from the minor lineage. Statistically significant differences in averaged sample values (to account for nonindependent intrahost sampling) are indicated by asterisks (**, $P < 0.005$; *, $P < 0.05$).

samples for group ST occupied the lower left quadrants for all three measures of the 1-step network components (Fig. 4A). Taken together, the data indicated that ST viromes were stably maintained for 16 weeks.

HVR1 evolution is markedly conservative in ST patient viromes. Partitioning of the conserved E1 region (195 bp) and HVR1 (81 bp) distinguished nonsynonymous mutation at HVR1 as the main determinant differentiating ST and AD phenotypes (Fig. 5B). The ratio of nonsynonymous to synonymous evolutionary changes (*dN*/*dS*) for the HVR1 portions of ST sequences indicated that HVR1 was predominantly under purifying selection pressures (Fig. 5F). Significant differences in age and the presence or absence of cirrhosis between ST and AD patients were also observed (*P* value = 0.017). All patients were chronically infected for at least 3 years (25). The parameters of age and cirrhosis were introduced as additional surrogate markers for the duration of infection, and HVR1 evolution was subsequently viewed in this context (33–35).

Separation of the patient cohort into ST and AD viromes was also evident from the divergence of the HVR1 pool from the initial samples (Fig. 6A) ($P < 0.01$). All patients identified as AD patients from *k*-step network analysis showed marked separation of the HVR1 epitope away from the original quasispecies. This was in contrast to ST patients, who demonstrated minimal movement of the HVR1 quasispecies over time. The dominance of individual HVR1 variants within ST viromes was fixed, with little evidence of epitope evolution (Fig. 6B). In contrast, the dominant HVR1 for AD group viromes was seen to change at multiple sites within 8 weeks (Fig. 6C). Nevertheless, the mutational capacities of HVR1 were similar for both ST3 and AD1. Sequence analysis across all unique ST3 and AD1 HVR1 variants isolated revealed that just 10/27 and 11/27 sites, respectively, were conserved (Fig. 6D).

Unique HVR1 variants from each sample set were subjected to 1-step network analysis as detailed previously for the nucleotide data, and variants linked by a single amino acid substitution were assigned to 1-step components. The entire HVR1 variant pool for ST1, ST3, ST4, and ST5 formed a single 1-step component, with the dominant HVR1 exhibiting the highest edge degree (Table 2

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

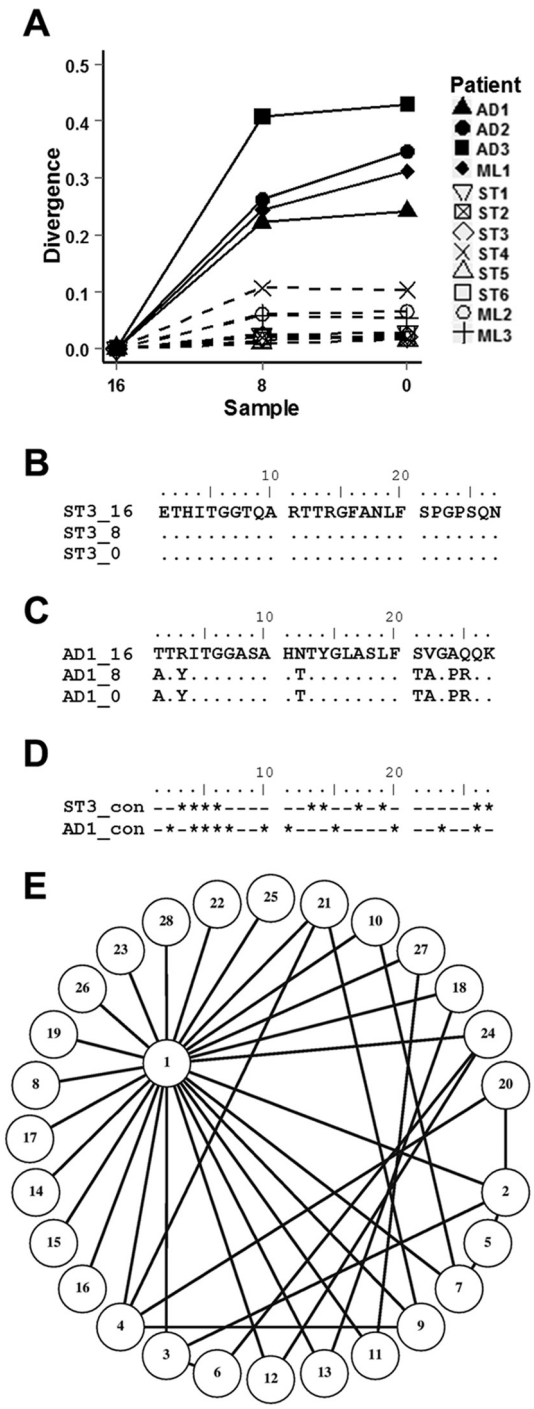


FIG 6 HVR1 evolution is conservative in ST group viromes. (A) HVR1 amino acid divergence from the original sample was calculated as the pairwise mean distance for ST patient viromes (dashed lines) and AD patient viromes (solid lines). The dominant ST3 HVR1 variant remained unchanged over the 16-week study period (B), whereas the AD1 HVR1 variant's dominance altered between samples, with multiple mutations across the 27-amino-acid length of the epitope (C). (D) ST3 and AD1 HVR1 sequence conservation across all isolates demonstrates that ST3 variants retain mutational capacity. Conserved sites are denoted by asterisks, and variant sites are denoted by dashes. (E) All 28 unique HVR1 variants isolated from ST3 had a single amino acid separation from at least one other HVR1 motif within the sample set, as shown in the 1-step network graph. Variants are numbered in order of the averaged frequency across all three samples (for variant 1, 88.3%; and for variant 28, 0.06%).

and Fig. 6E). Remarkably, for ST5, only two unique HVR1 variants (occupying 99.9% and 0.1% of the sample space) were recovered from three independent sample preparations and >10,000 individual reads sequenced. For ST2 and ST6, across all three sampling points combined, the dominant HVR1 1-step component accounted for 96.9% and 99.8% of the sample space, respectively. In contrast, the dominant HVR1 1-step components for AD1, AD2, and AD3 accounted for 62.4%, 33.3%, and 49.1% of the sample space, respectively.

AD HVR1 variant pools demonstrate pronounced physiochemical changes. HVR1 microdomains participating in SRBI interactions, influencing infectivity and encompassing a neutralizing epitope, have been defined for the H77 HVR1 variant (16). We sought to map the observed HVR1 mutations within our data to these sites (Fig. 7).

As anticipated, conservative changes were observed for the majority of ST HVR1 sites. The most noticeable exception was for ST4, which had the largest recorded HVR1 divergence of the ST group (Fig. 6A). A considerable proportion of this change occurred within the SRBI interacting microdomain (Fig. 7, residues 384 to 396). AD viromes, with the exception of AD1, exhibited high diversity within the proposed nAb epitope for H77 and residues 397 and 398, which are linked to infectivity (Fig. 7) (16). Additionally, the changes were for nonconservative amino acids. Within this domain, residues 403, 406, and 407 emerged as highly conserved across all viromes, which is suggestive of a discrete preservation of function. Overall, the pattern of diverse mutational change seen for AD sequences implicates nAb targeting and the modulation of processes governing infectivity.

IgG binding of virions is associated with the dominant HVR1 amino acid epitope. We previously showed that IgG-bound virions can be fractionated away from IgG-free virions through affinity chromatography (22, 26, 36). In this study, the HVR1 profile of IgG-bound virions in the 0-week samples was determined, and the prevalence of the IgG-associated HVR1 motifs over the 16-week sampling period was evaluated retrospectively.

Ten of the 12 patients were positive for HCV RNA following IgG fractionation (ST5 and AD2 were identified as negative for HCV RNA in the IgG-bound sample). In cases where the predicted HVR1 of IgG-bound virions initially occupied >80% of the sample space, it remained so for the subsequent samples. In cases where the predicted HVR1 of IgG-bound virions was initially between 0 and 40%, the occupation of the sample space expanded by an additional 25 to 95% (Fig. 8A). The homogeneity of ST group samples was clear for ST1, ST2, and ST3, as a single HVR1 amino acid motif accounted for >80% of the total virome across all three sampling points. IgG binding was associated with virus sequences that coded for the 0-week dominant HVR1 variant in ST4, ST6, and ML2 (Fig. 8A). Within the latter three viromes, the between-sample dominant HVR1 variants differed by a single amino acid substitution event (data not shown). With respect to ML2, the HVR1 variant identified following IgG fractionation formed part of the minor lineage variant pool (Table 2).

HVR1 epitope heterogeneity was observed in just 2 of the 10 IgG-bound virus RNA-positive samples, both of which had AD group viromes (AD3 and ML1, respectively). For AD1 and AD3, the predicted HVR1 of IgG-bound virions was present at <2% of the 16-week sample space and rose to >90% in both subsequent samples (Fig. 8A). Given the positive selection of HVR1 in this group (Fig. 5F) and the pronounced nonconservative amino acid

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

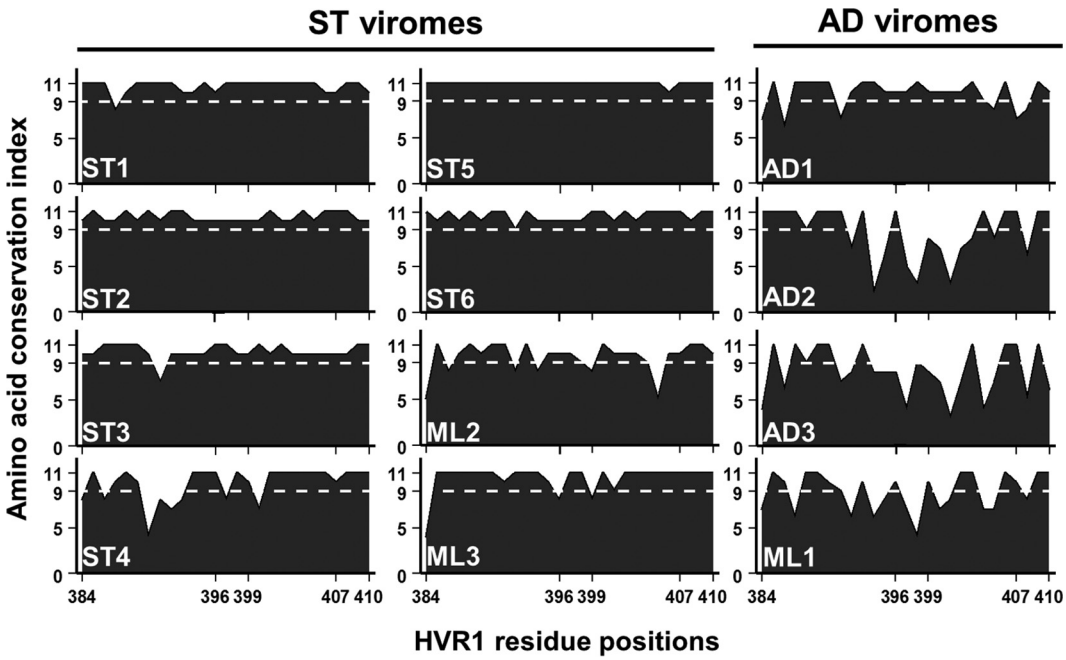


FIG 7 Amino acid conservation of patient-specific HVR1 variant pools is pronounced among ST viromes. An amino acid physiochemical conservation plot was prepared for each of the 12 patients, based on the AMAS method of multiple-sequence alignment, and scored (*y* axis). Values of >9 reflect ultraconservative changes or absolute conservation (dashed white lines). Functional microdomains as characterized by Guan et al. (16) for the H77 HVR1, which are purported to participate in SRBI interactions (residues 384 to 396) and to contain a nAb epitope (residues 399 to 407), are identified (*x* axis), while residues 397 and 398 and 408 to 410 modulate infectivity. HVR1 residues are numbered in accordance to the H77 reference genome.

substitutions within the putative nAb epitope (Fig. 7A, residues 399 to 407), the data support the hypothesis that, in both instances, these variants were subject to humoral immune targeting. This conclusion was further strengthened by the isolation of five unique AD3 HVR1 epitopes following IgG fractionation, three of which were detectable in the UDPS data (Fig. 8B). The sum of these data is that the AD3 HVR1 profile was indicative of antibody targeting and removal of variants, which reflected the between-sample dominant HVR1 profile.

While effective IgG binding of virions occurred for both the ST and AD phenotypes, only the AD3 HVR1 profile was indicative of neutralization during the 16-week time frame examined here. We previously reported that the predicted HVR1 of IgG-bound virions was associated with the collapse of the constituent virion population in a process that was measured in years rather than weeks or months (22). Consequently, the 16-week study period is likely to be insufficient to determine the full extent of neutralization efficacy of the antibody response. However, the antibody repertoire does appear to be capable of readily recognizing virions as new HVR1 epitopes emerge from the background variant pool, regardless of HVR1 phenotype (Fig. 8).

DISCUSSION

Visualization of the data through the generation of *k*-step networks and analysis of the constituent 1-step network components revealed viromes governed by either stably or temporally dominant master sequences. We categorized these phenotypes as ST or AD based on the evident evolutionary divergence within the sequence sets. ST virome HVR1 variant pools converged around the dominant HVR1 epitope, with most variants separated from the dominant epitope by a single amino acid substitution. IgG binding of virions was associated with the sample-specific dominant HVR1 in both ST and AD groups but was indicative of neutralization only for AD3 variants during the study period.

Despite a large volume of research documenting the genetic variability of HVR1, discrete windows of intrahost evolution in chronic infection are lacking in the literature. Initial clonal analysis of samples, collected at 2-week intervals over 16 weeks, identified both divergent and stationary HVR1 evolutionary phenotypes (25). In the present study, UDPS was utilized to scrutinize this window of HVR1 evolution in treatment-naive patients, all of whom were chronically infected with HCV for at least 3 years. Next-generation sequencing technologies facilitate a more complete mining of the quasispecies pool. Our analysis concurs with estimates that variants present at a frequency of >5% of the sample space are reliably detectable using clonal methods (37, 38). Nevertheless, ST and AD phenotype classification was not achievable through clonal analysis, as the true quasispecies spectrum was masked by a few dominant sequences (Table 2).

For all 12 patients, the sample-specific viromes were organized around dominant haplotypes or master sequences that were highly connected to the quasispecies pool, largely through single-point synonymous substitutions (Fig. 2 and 3). ST viromes were relatively homogeneous and explored a narrow sequence space. Additionally, significant associations of ST viromes were made with older patients and patients with cirrhosis (Table 1). While sequence heterogeneity and rapid virus evolution have been associated with the initial development of fibrosis, the homogeneous viromes seen here may be a reflection of further advancement of liver disease (39–41). Indeed, HCV viromes with low-diversity HVR1 variant pools have, over time, been linked to increasing disease severity (42).

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

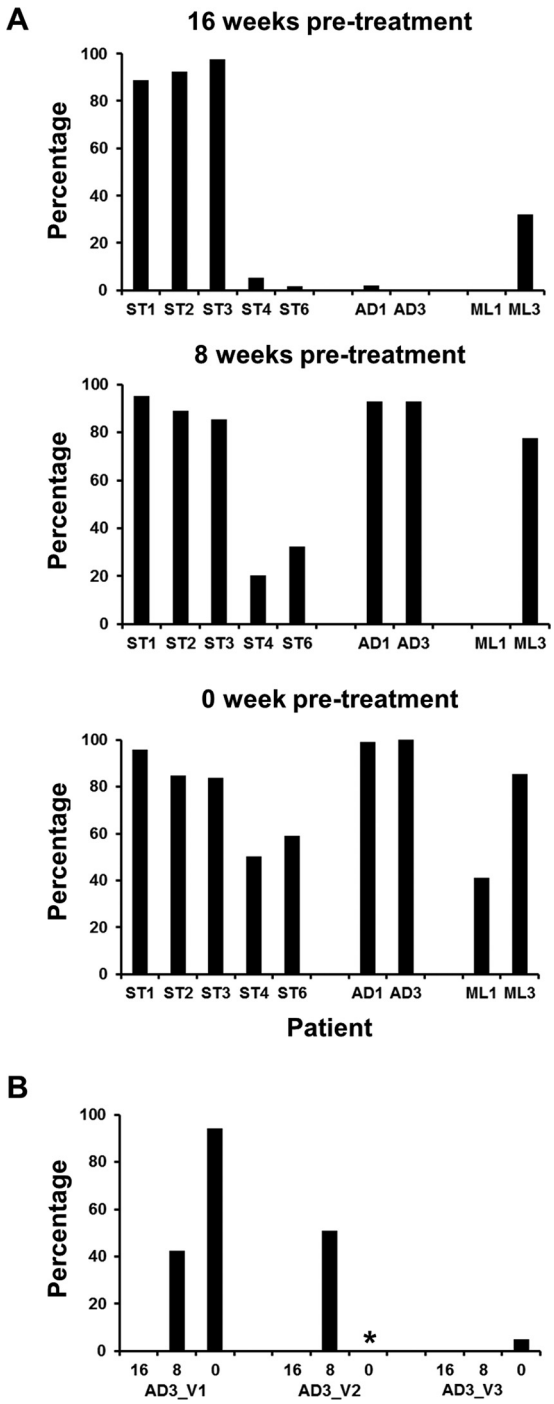


FIG 8 IgG binding of virions was associated with the dominant HVR1 in the 0-week sample. All samples taken on the day of treatment were subjected to fractionation of IgG-bound virions followed by extraction of virus genomes and determination of the predicted HVR1 epitopes. (A) UDPS data from pretreatment samples (16 and 8 weeks prior) were retrospectively reviewed to assess the prevalence of HVR1 epitopes associated with IgG-bound virions. The *x* axis gives the patient identifier, and the *y* axis gives the percentage of the sample space occupied by sequences containing the predicted IgG-bound HVR1. (B) Three AD3 HVR1 variants associated with IgG-bound virions (AD3_V1, accession number KT193831; AD3_V2, accession number KT193830; and AD3_V3, accession number KT193829) were detectable in the UDPS data. Partition analysis by variant revealed that IgG binding was associated with variant emergence and removal. Specifically, the AD3_V2 epitope was isolated only from the IgG-bound fraction, not from UDPS analysis of whole serum, for the 0-week sample (asterisk). In the 8-week sample, this variant accounted for 50.8% of the UDPS sample space. The *y* axis gives the percentage of sample space occupation of the predicted IgG-bound HVR1, and the *x* axis defines the IgG-bound variant occupation of the sample space by time point.

We report significant differences between ST and AD group haplotype profiles solely in the exploration of nonsynonymous sequence space across HVR1 and the subsequent divergence from the initial sample (Fig. 5 and 6). The intriguing observation with respect to ST HVR1 variant pools was the interconnectedness of the dominant HVR1 variant to remaining minor variants, largely through single amino acid point mutations (Fig. 6E). We previously detailed an HCV genotype 4a mixed-lineage infection in which a low-diversity minor lineage expanded into the sample space over a period of 10 years, to dominate the virome *in toto* during the last 2 years (percentage sample space minimum, 0.4%; and maximum, 96.9%) (22). A conservative pattern of synonymous mutation was observed that parallels the ST phenotypes reported here, with the distinction that the HVR1 variant dominated the sample space for years rather than weeks.

We did not observe differences between ST and AD patients with respect to IgG binding, and we cannot exclude the possibility that epitopes outside the N terminus of E2 may exhibit immunodominance. However, significant differences distinguished humoral immune targeting of the HVR1 ST and AD variant pools. AD HVR1 epitopes were under strong positive selection pressures and exhibited frequent variant replacement rather than sequence diversification (Fig. 5). Furthermore, the predicted HVR1 profile of AD3 IgG-bound virions argues in favor of direct humoral immune targeting, and this is strengthened by the observation of pronounced nonconservative mutational changes in the putative nAb epitope of AD HVR1 variants (Fig. 7 and 8B) (16).

Lower levels of nucleotide substitution in HVR1 have been reported for patients with hypogammaglobulinemia, suggesting that sequence evolution in this region of the genome is primarily linked to humoral immune pressure (43–45). For each patient with detectable IgG binding of virions, the associated HVR1 variant was dominant and centrally placed within the virome network (Fig. 6E), which has previously been associated with greater cross-immunoreactivity (46). However, given the short time frame and the commencement of treatment following the completion of this study, we were unable to confirm whether the HVR1 profile of detectable virus-antibody interactions seen for AD3 extended to the wider sample set.

ST viromes additionally displayed considerable global stability (Fig. 4). Original antigenic sin, the preferential activation of immune memory against a similar yet nonidentical antigen following a reinfection event, has been described for HCV, dengue virus, and influenza virus (47–49). The delayed removal of the dominant variant targeted by a weakened nAb response is known to extend to years (22, 50). In the context of a convergent HVR1 variant pool, the presentation of successive, antigenically similar but nonidentical epitopes (Fig. 6E) may impart a cumulative weakening of the nAb response required for effective virion neutralization (51–53). Consequently, the minimum binding threshold required for virion neutralization is more difficult to achieve, which concomitantly facilitates the maintenance of related minor variants (54, 55). Together, the data support a model of antigenic cooperation enabled in viromes, organized around a single dominant variant (51).

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

The persistent dominance of specific HVR1 epitopes indicates diminished humoral immune pressures. Indeed, stronger nAb cross-reactivity with historic than with current HVR1 epitopes has been shown (50). Alleviation of potentially deleterious nonsynonymous mutations forced upon this region allows for enhancement of fitness through exploration of the synonymous sequence space (56). In the context of an error-prone polymerase and a high replication rate, the maintenance of singly dominant HVR1 epitopes over weeks (as observed here) and years is indicative of host adaptation and/or the maintenance of functional advantages (21, 22, 57).

The acquisition of host-specific adaptions to HVR1 has the potential to enhance infectivity and receptor recognition (16, 58–60). As the within-host period of infection extends, we predict a preferential evolution toward ST rather than continued rapid and nonconservative epitope change. Based on our observations over 16 weeks, AD1 is moving toward an ST group phenotype given the contraction of the network to a single 1-step component in the 0-week sample (Fig. 2 and 3). We note that the adaptive capacity of HVR1 is retained regardless of the specific mutational phenotype (Fig. 6D), and we recognize that reversion between phenotypes over time is feasible, depending on the within-host environment (57).

In summary, homogeneous HVR1 populations arise as a consequence of long-term, host-specific, pervasive humoral immune selection. Complex viromes reflect a population dynamic that explores a more expansive sequence space in an attempt to find within-host fitness optima. The benefits of nearly clonal HVR1 dominance in the virus are ill defined, but this phenomenon does imply adaptation of HCV to its host.

ACKNOWLEDGMENTS

Daniel Schmidt-Martin was funded by Molecular Medicine Ireland through the Clinician Scientist Structured Training Programme.

We thank David S. Campo (Division of Viral Hepatitis, Centers for Disease Control and Prevention) for his assistance with 1-step and *k*-step network visualization.

REFERENCES

1. **Yamane D, McGivern DR, Masaki T, Lemon SM.** 2013. Liver injury and disease pathogenesis in chronic hepatitis C. Curr Top Microbiol Immunol **369:**263–288. http://dx.doi.org/10.1007/978-3-642-27340-7_11.
2. **Gower E, Estes C, Blach S, Razavi-Shearer K, Razavi H.** 2014. Global epidemiology and genotype distribution of the hepatitis C virus infection. J Hepatol **61:**S45–S57. http://dx.doi.org/10.1016/j.jhep.2014.07.027.
3. **Santantonio T, Wiegand J, Gerlach JT.** 2008. Acute hepatitis C: current status and remaining challenges. J Hepatol **49:**625–633. http://dx.doi.org/10.1016/j.jhep.2008.07.005.
4. **Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science **244:**359–362. http://dx.doi.org/10.1126/science.2523562.
5. **Smith BD, Morgan RL, Beckett GA, Falck-Ytter Y, Holtzman D, Ward JW.** 2012. Hepatitis C virus testing of persons born during 1945–1965: recommendations from the Centers for Disease Control and Prevention. Ann Intern Med **157:**817–822. http://dx.doi.org/10.7326/0003-4819-157-9-201211060-00529.
6. **Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P.** 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology **59:**318–327. http://dx.doi.org/10.1002/hep.26744.
7. **Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, Barnes E.** 2015. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology **61:**77–87. http://dx.doi.org/10.1002/hep.27259.
8. **Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, Crawford K, Bonino F, Saracco G, Choo QL, Houghton M, Han JH.** 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:**842–848. http://dx.doi.org/10.1016/0042-6822(91)90104-J.
9. **Khan AG, Whidby J, Miller MT, Scarborough H, Zatorski AV, Cygan A, Price AA, Yost SA, Bohannon CD, Jacob J, Grakoui A, Marcotrigiano J.** 2014. Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. Nature **509:**381–384. http://dx.doi.org/10.1038/nature13117.
10. **Dowd KA, Netski DM, Wang XH, Cox AL, Ray SC.** 2009. Selection pressure from neutralizing antibodies drives sequence evolution during acute infection with hepatitis C virus. Gastroenterology **136:**2377–2386. http://dx.doi.org/10.1053/j.gastro.2009.02.080.
11. **Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzera A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ.** 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science **288:**339–344. http://dx.doi.org/10.1126/science.288.5464.339.
12. **Bartosch B, Verney G, Dreux M, Donot P, Morice Y, Penin F, Pawlotsky JM, Lavillette D, Cosset FL.** 2005. An interplay between hypervariable region 1 of the hepatitis C virus E2 glycoprotein, the scavenger receptor BI, and high-density lipoprotein promotes both enhancement of infection and protection against neutralizing antibodies. J Virol **79:**8217–8229. http://dx.doi.org/10.1128/JVI.79.13.8217-8229.2005.
13. **Dao Thi VL, Granier C, Zeisel MB, Guerin M, Mancip J, Granio O, Penin F, Lavillette D, Bartenschlager R, Baumert TF, Cosset FL, Dreux M.** 2012. Characterization of hepatitis C virus particle subpopulations reveals multiple usage of the scavenger receptor BI for entry steps. J Biol Chem **287:**31242–31257. http://dx.doi.org/10.1074/jbc.M112.365924.
14. **Wahid A, Dubuisson J.** 2013. Virus-neutralizing antibodies to hepatitis C virus. J Viral Hepat **20:**369–376. http://dx.doi.org/10.1111/jvh.12094.
15. **Farci P, Shimoda A, Wong D, Cabezon T, De Gioannis D, Strazzera A, Shimizu Y, Shapiro M, Alter HJ, Purcell RH.** 1996. Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. Proc Natl Acad Sci U S A **93:**15394–15399. http://dx.doi.org/10.1073/pnas.93.26.15394.
16. **Guan M, Wang W, Liu X, Tong Y, Liu Y, Ren H, Zhu S, Dubuisson J, Baumert TF, Zhu Y, Peng H, Aurelian L, Zhao P, Qi Z.** 2012. Three different functional microdomains in the hepatitis C virus hypervariable region 1 (HVR1) mediate entry and immune evasion. J Biol Chem **287:**35631–35645. http://dx.doi.org/10.1074/jbc.M112.382341.
17. **Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Ohkoshi S, Shimotohno K.** 1991. Hypervariable regions in the putative glycoprotein of hepatitis C virus. Biochem Biophys Res Commun **175:**220–228. http://dx.doi.org/10.1016/S0006-291X(05)81223-9.
18. **Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, Cameron B, Maher L, Dore GJ, White PA, Lloyd AR.** 2011. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. PLoS Pathog **7:**e1002243. http://dx.doi.org/10.1371/journal.ppat.1002243.
19. **Elena SF, Sole RV, Sardanyes J.** 2010. Simple genomes, complex interactions: epistasis in RNA virus. Chaos **20:**026106. http://dx.doi.org/10.1063/1.3449300.
20. **Gismondi MI, Díaz Carrasco JM, Valva P, Becker PD, Guzmán CA, Campos RH, Preciado MV.** 2013. Dynamic changes in viral population structure and compartmentalization during chronic hepatitis C virus infection in children. Virology **447:**187–196. http://dx.doi.org/10.1016/j.virol.2013.09.002.
21. **Ramachandran S, Campo DS, Dimitrova ZE, Xia GL, Purdy MA, Khudyakov YE.** 2011. Temporal variations in the hepatitis C virus intrahost population during chronic infection. J Virol **85:**6369–6380. http://dx.doi.org/10.1128/JVI.02204-10.
22. **Palmer BA, Dimitrova Z, Skums P, Crosbie O, Kenny-Walsh E, Fanning LJ.** 2014. Analysis of the evolution and structure of a complex intrahost viral population in chronic hepatitis C virus mapped by ultradeep pyrosequencing. J Virol **88:**13709–13721. http://dx.doi.org/10.1128/JVI.01732-14.
23. **Li H, Hughes AL, Bano N, McArdle S, Livingston S, Deubner H, McMahon BJ, Townshend-Bulson L, McMahan R, Rosen HR, Gretch DR.** 2011. Genetic diversity of near genome-wide hepatitis C virus sequences during chronic infection: evidence for protein structural conservation over time. PLoS One **6:**e19562. http://dx.doi.org/10.1371/journal.pone.0019562.

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

24. **Lu L, Tatsunori N, Li C, Waheed S, Gao F, Robertson BH.** 2008. HCV selection and HVR1 evolution in a chimpanzee chronically infected with HCV-1 over 12 years. Hepatol Res **38:**704–716. http://dx.doi.org/10.1111/j.1872-034X.2008.00320.x.
25. **Schmidt-Martin D, Crosbie O, Kenny-Walsh E, Fanning LJ.** 2015. Intensive temporal mapping of hepatitis C hypervariable region 1 quasispecies provides novel insights into hepatitis C virus evolution in chronic infection. J Gen Virol **96:**2145–2156. http://dx.doi.org/10.1099/vir.0.000149.
26. **Palmer BA, Moreau I, Levis J, Harty C, Crosbie O, Kenny-Walsh E, Fanning LJ.** 2012. Insertion and recombination events at hypervariable region 1 over 9.6 years of hepatitis C virus chronic infection. J Gen Virol **93:**2614–2624. http://dx.doi.org/10.1099/vir.0.045344-0.
27. **Skums P, Dimitrova Z, Campo DS, Vaughan G, Rossi L, Forbi JC, Yokosawa J, Zelikovsky A, Khudyakov Y.** 2012. Efficient error correction for next-generation sequencing of viral amplicons. BMC Bioinformatics **13**(Suppl 10):S6. http://dx.doi.org/10.1186/1471-2105-13-S10-S6.
28. **Campo DS, Dimitrova Z, Yamasaki L, Skums P, Lau DT, Vaughan G, Forbi JC, Teo CG, Khudyakov Y.** 2014. Next-generation sequencing reveals large connected networks of intra-host HCV variants. BMC Genomics **15**(Suppl 5):S4. http://dx.doi.org/10.1186/1471-2164-15-S5-S4.
29. **Batagelj V, Mrvar A.** 2004. Pajek—analysis and visualization of large networks, p 77–103. *In* Jünger M, Mutzel P (ed), Graph drawing software. Springer, Berlin, Germany.
30. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725–2729. http://dx.doi.org/10.1093/molbev/mst197.
31. **Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ.** 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics **25:**1189–1191. http://dx.doi.org/10.1093/bioinformatics/btp033.
32. **Livingstone CD, Barton GJ.** 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci **9:**745–756.
33. **Poynard T, Bedossa P, Opolon P.** 1997. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC Groups. Lancet **349:**825–832.
34. **Hajarizadeh B, Grebely J, Dore GJ.** 2013. Epidemiology and natural history of HCV infection. Nat Rev Gastroenterol Hepatol **10:**553–562. http://dx.doi.org/10.1038/nrgastro.2013.107.
35. **Pradat P, Voirin N, Tillmann HL, Chevallier M, Trepo C.** 2007. Progression to cirrhosis in hepatitis C patients: an age-dependent process. Liver Int **27:**335–339. http://dx.doi.org/10.1111/j.1478-3231.2006.01430.x.
36. **Moreau I, O'Sullivan H, Murray C, Levis J, Crosbie O, Kenny-Walsh E, Fanning LJ.** 2008. Separation of hepatitis C genotype 4a into IgG-depleted and IgG-enriched fractions reveals a unique quasispecies profile. Virol J **5:**103. http://dx.doi.org/10.1186/1743-422X-5-103.
37. **Polyak SJ, Faulkner G, Carithers RL, Jr, Corey L, Gretch DR.** 1997. Assessment of hepatitis C virus quasispecies heterogeneity by gel shift analysis: correlation with response to interferon therapy. J Infect Dis **175:**1101–1107. http://dx.doi.org/10.1086/516448.
38. **McCaughan GW, Laskus T, Vargas HE.** 2003. Hepatitis C virus quasispecies: misunderstood and mistreated? Liver Transpl **9:**1048–1052. http://dx.doi.org/10.1053/jlts.2003.50260.
39. **Domingo E, Sheldon J, Perales C.** 2012. Viral quasispecies evolution. Microbiol Mol Biol Rev **76:**159–216. http://dx.doi.org/10.1128/MMBR.05023-11.
40. **Lauring AS, Andino R.** 2010. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog **6:**e1001005. http://dx.doi.org/10.1371/journal.ppat.1001005.
41. **Wang XH, Netski DM, Astemborski J, Mehta SH, Torbenson MS, Thomas DL, Ray SC.** 2007. Progression of fibrosis during chronic hepatitis C is associated with rapid virus evolution. J Virol **81:**6513–6522. http://dx.doi.org/10.1128/JVI.02276-06.
42. **Sullivan DG, Bruden D, Deubner H, McArdle S, Chung M, Christensen C, Hennessy T, Homan C, Williams J, McMahon BJ, Gretch DR.** 2007. Hepatitis C virus dynamics during natural infection are associated with long-term histological outcome of chronic hepatitis C disease. J Infect Dis **196:**239–248. http://dx.doi.org/10.1086/518895.
43. **Booth JC, Kumar U, Webster D, Monjardino J, Thomas HC.** 1998. Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. Hepatology **27:**223–227. http://dx.doi.org/10.1002/hep.510270134.
44. **Landau DA, Saadoun D, Calabrese LH, Cacoub P.** 2007. The pathophysiology of HCV induced B-cell clonal disorders. Autoimmun Rev **6:**581–587. http://dx.doi.org/10.1016/j.autrev.2007.03.010.
45. **Gisbert JP, Garcia-Buey L, Pajares JM, Moreno-Otero R.** 2005. Systematic review: regression of lymphoproliferative disorders after treatment for hepatitis C infection. Aliment Pharmacol Ther **21:**653–662. http://dx.doi.org/10.1111/j.1365-2036.2005.02395.x.
46. **Campo DS, Dimitrova Z, Yokosawa J, Hoang D, Perez NO, Ramachandran S, Khudyakov Y.** 2012. Hepatitis C virus antigenic convergence. Sci Rep **2:**267. http://dx.doi.org/10.1038/srep00267.
47. **Proust B, Dubois F, Bacq Y, Le Pogam S, Rogez S, Levillain R, Goudeau A.** 2000. Two successive hepatitis C virus infections in an intravenous drug user. J Clin Microbiol **38:**3125–3127.
48. **Kim JH, Skountzou I, Compans R, Jacob J.** 2009. Original antigenic sin responses to influenza viruses. J Immunol **183:**3294–3301. http://dx.doi.org/10.4049/jimmunol.0900398.
49. **Midgley CM, Bajwa-Joseph M, Vasanawathana S, Limpitikul W, Wills B, Flanagan A, Waiyaiya E, Tran HB, Cowper AE, Chotiyarnwong P, Grimes JM, Yoksan S, Malasit P, Simmons CP, Mongkolsapaya J, Screaton GR.** 2011. An in-depth analysis of original antigenic sin in dengue virus infection. J Virol **85:**410–421. http://dx.doi.org/10.1128/JVI.01826-10.
50. **von Hahn T, Yoon JC, Alter H, Rice CM, Rehermann B, Balfe P, Mc Keating JA.** 2007. Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo. Gastroenterology **132:**667–678. http://dx.doi.org/10.1053/j.gastro.2006.12.008.
51. **Skums P, Bunimovich L, Khudyakov Y.** 2015. Antigenic cooperation among intrahost HCV variants organized into a complex network of cross-immunoreactivity. Proc Natl Acad Sci U S A **112:**6653–6658. http://dx.doi.org/10.1073/pnas.1422942112.
52. **Parsons MS, Muller S, Kohler H, Grant MD, Bernard NF.** 2013. On the benefits of sin: can greater understanding of the 1F7-idiotypic repertoire freeze enhance HIV vaccine development? Hum Vaccin Immunother **9:**1532–1538. http://dx.doi.org/10.4161/hv.24460.
53. **Bartosch B, Bukh J, Meunier JC, Granier C, Engle RE, Blackwelder WC, Emerson SU, Cosset FL, Purcell RH.** 2003. In vitro assay for neutralizing antibody to hepatitis C virus: evidence for broadly conserved neutralization epitopes. Proc Natl Acad Sci U S A **100:**14199–14204. http://dx.doi.org/10.1073/pnas.2335981100.
54. **Meyer K, Banerjee A, Frey SE, Belshe RB, Ray R.** 2011. A weak neutralizing antibody response to hepatitis C virus envelope glycoprotein enhances virus infection. PLoS One **6:**e23699. http://dx.doi.org/10.1371/journal.pone.0023699.
55. **Dowd KA, Pierson TC.** 2011. Antibody-mediated neutralization of flaviviruses: a reductionist view. Virology **411:**306–315. http://dx.doi.org/10.1016/j.virol.2010.12.020.
56. **Cuevas JM, Domingo-Calap P, Sanjuan R.** 2012. The fitness effects of synonymous mutations in DNA and RNA viruses. Mol Biol Evol **29:**17–20. http://dx.doi.org/10.1093/molbev/msr179.
57. **Farci P, Wollenberg K, Diaz G, Engle RE, Lai ME, Klenerman P, Purcell RH, Pybus OG, Alter HJ.** 2012. Profibrogenic chemokines and viral evolution predict rapid progression of hepatitis C to cirrhosis. Proc Natl Acad Sci U S A **109:**14562–14567. http://dx.doi.org/10.1073/pnas.1210592109.
58. **Hsu M, Zhang J, Flint M, Logvinoff C, Cheng-Mayer C, Rice CM, McKeating JA.** 2003. Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. Proc Natl Acad Sci U S A **100:**7271–7276. http://dx.doi.org/10.1073/pnas.0832180100.
59. **Bankwitz D, Steinmann E, Bitzegeio J, Ciesek S, Friesland M, Herrmann E, Zeisel MB, Baumert TF, Keck ZY, Foung SK, Pecheur EI, Pietschmann T.** 2010. Hepatitis C virus hypervariable region 1 modulates receptor interactions, conceals the CD81 binding site, and protects conserved neutralizing epitopes. J Virol **84:**5751–5763. http://dx.doi.org/10.1128/JVI.02200-09.
60. **Prentoe J, Jensen TB, Meuleman P, Serre SB, Scheel TK, Leroux-Roels G, Gottwein JM, Bukh J.** 2011. Hypervariable region 1 differentially impacts viability of hepatitis C virus strains of genotypes 1 to 6 and impairs virus neutralization. J Virol **85:**2224–2234. http://dx.doi.org/10.1128/JVI.01594-10.

Downloaded from http://jvi.asm.org/ on September 1, 2017 by IRIS

Chapter 8

Analysis of IgG binding patterns and influence on HVR1 sequence change

8.1 Introduction

We have, using cloning strategies identified varying patterns of HCV hypervariable 1 (HVR1) region change among 23 subjects with chronic infection, who were prospectively studied over a 16 week period in advance of commencing treatment(302).

A number of previous studies have used co immunoprecipitation and, more latterly, immunoglobulin separation spin columns to evaluate to contribution of adaptive antibody mediated immune responses to the emergence of new HCV quasispecies variants(283, 410, 411).

In order to identify antibody driven change, we separated the serum into Immunoglobulin G (IgG) enriched and IgG depleted fractions and, amplified and sequenced HVR1 in the respective fractions. Using preliminary data from our prospective study of temporal change in HVR1 quasispecies, we selected a group of subjects with differing patterns of QS change. The subjects are labelled in accordance with previous chapters.

8.2 Methods

See Chapter 2 section 2.7.

8.3 Results

Temporal change in quasispecies parameters.

	Genotype	Cirrhosis	SVR	Phylogenetic change	Time ordered phylogeny	Single Dominant Subpopulation	Number of Subpopulations	Replacement of Dominant Subpopulation	Selection
A	1b	Y	Y	N		+	2		
B	1b	Y	N	N		+	3		
C	1b	Y	N	N		+	1		
G	3a	N	Y	Y		-	10	+	+
H	3a	N	Y	Y	+	-	8	+	
L	3a	N	Y	Y		-	2	+	
N	3a	N	Y	N		+	3		
Q	3a	N	Y	Y	+	-	7	+	+
T	3a	Y	Y	N		+	1		

Table 1.

Summary of results of the temporal change in HVR1 as described in the prospective cloning study. The samples chosen for immunoglobulin separation included a mix of genotypes and severity of underlying liver disease. The subjects chosen included those with time order phylogeny, those with a single subpopulation throughout the study period, subjects with a change in the dominant subpopulation, and subjects with evidence of sequence wide positive selection using PARRIS analysis www.datamonkey.org.

Using the first (week 16) and last (week 0) sample for each individual, we separated the virus, contained within 25µL aliquots of serum, from 9 subjects into immunoglobulin G bound and immunoglobulin G free fractions to evaluate for temporal changes in IgG binding over the 16 week study period.

The IgG free and bound fractions were then compared with the results of the prospective cloning study in order to evaluate for patterns of binding between subjects characterised by significant changes in the clones and those where there was no change (Table 1).

8.3.1 Subjects with no HVR1 change

8.3.1.1 Subject A

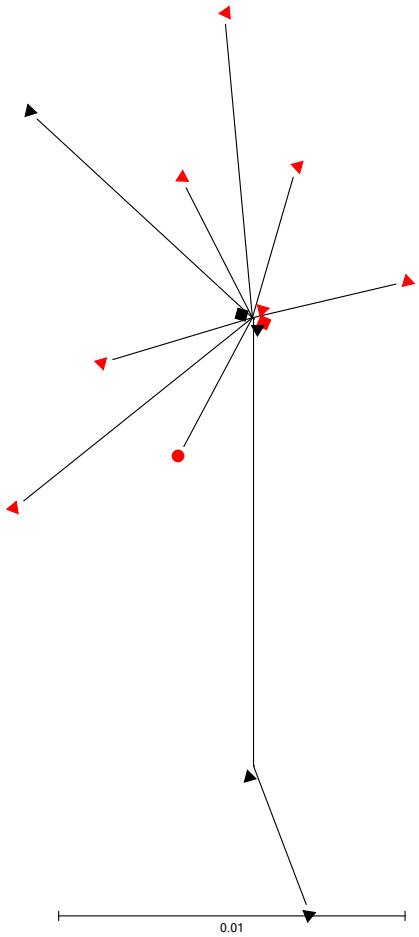


Fig 8.1. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pretreatment sample (red triangles) with the immunoglobulin bound (black square – Week 16/red square – Week 0) and free (black circle – Week 16/red circle – Week 0) sequences included for Subject A. The tree was generated using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5.

Patient A is characterised by a minimally evolving HVR1 sequence in the cloning study which is remarkably homogenous with only 3 of the 40 clones examined demonstrating any difference from the master sequences, and each of these have a single amino acid substitution within the HVR1. Subject A demonstrates IgG binding of the master sequence at both sampling points but, without any discernable influence on the QS pattern. Interestingly, at week 0 there appears to be no IgG free fraction, while at week 16 the IgG free fraction is characterised by a single amino acid substitution within the HVR1 (Fig. 8.1) when compared with the bound fraction.

8.3.1.2 Subject B

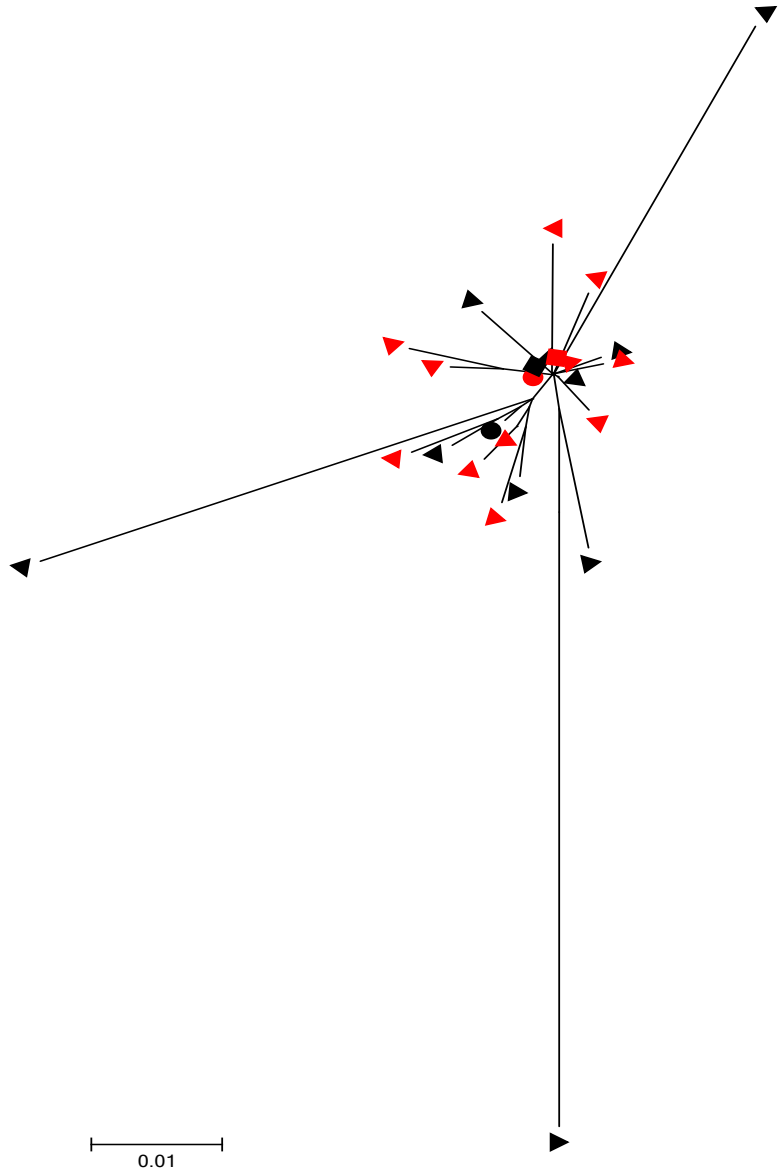


Fig 8.2. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pretreatment sample (red triangles) with the immunoglobulin bound (black square – Week 16/red square – Week 0) and free (black circle – Week 16/red circle – Week 0) sequences included for Subject B. The tree was generated using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5. Indentical nucleotide sequences overlap.

In subject B, the separation of the quasispecies into IgG enriched and depleted fractions resulted in the same HVR1 sequence in all fractions. This suggests either ineffective neutralisation or remote binding at another epitope (Fig 8.2.).

8.3.1.3 Subject C

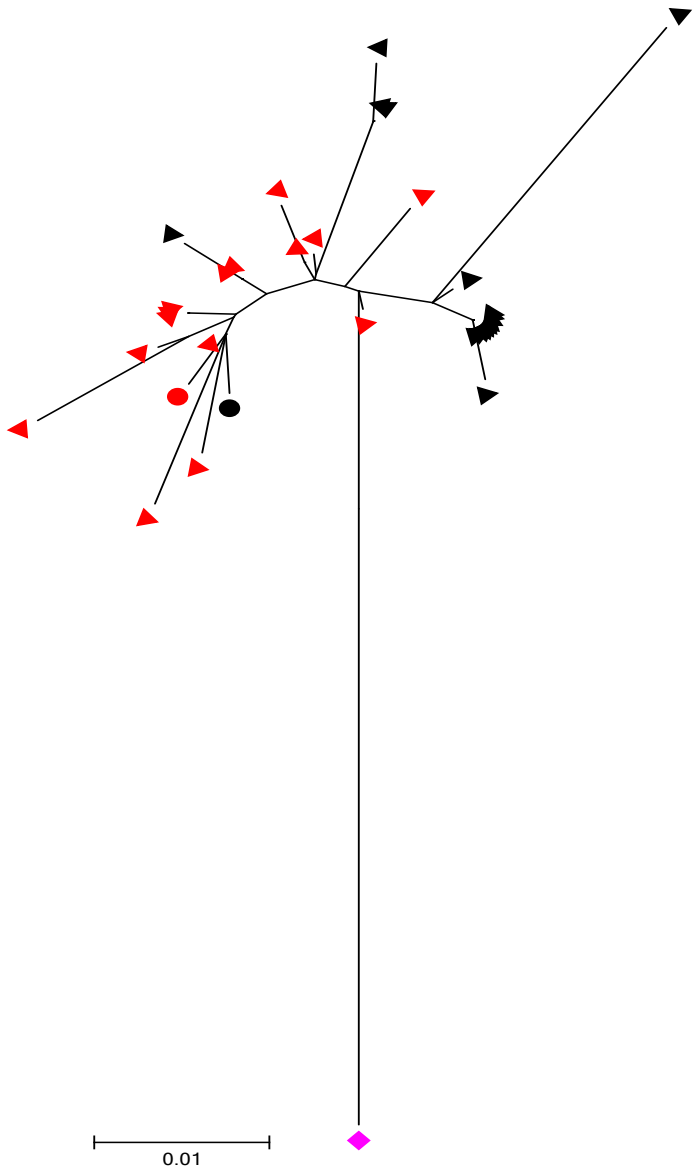


Fig 8.3. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pretreatment sample (red triangles), with the immunoglobulin bound (black square – Week 16/red square – Week 0) and free (black circle – Week 16/red circle – Week 0) sequences included for Subject C. The tree was generated using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5. The sequence designated with a pink diamond represents the original Anti D sequence which caused the iatrogenic infection in 1977.

Patient C demonstrated no HVR1 IgG enriched fragment. Both IgG free fractions correspond with the consensus master sequence which remained unchanged throughout the 16 week study period (Fig 8.3).

8.3.1.4Subject N

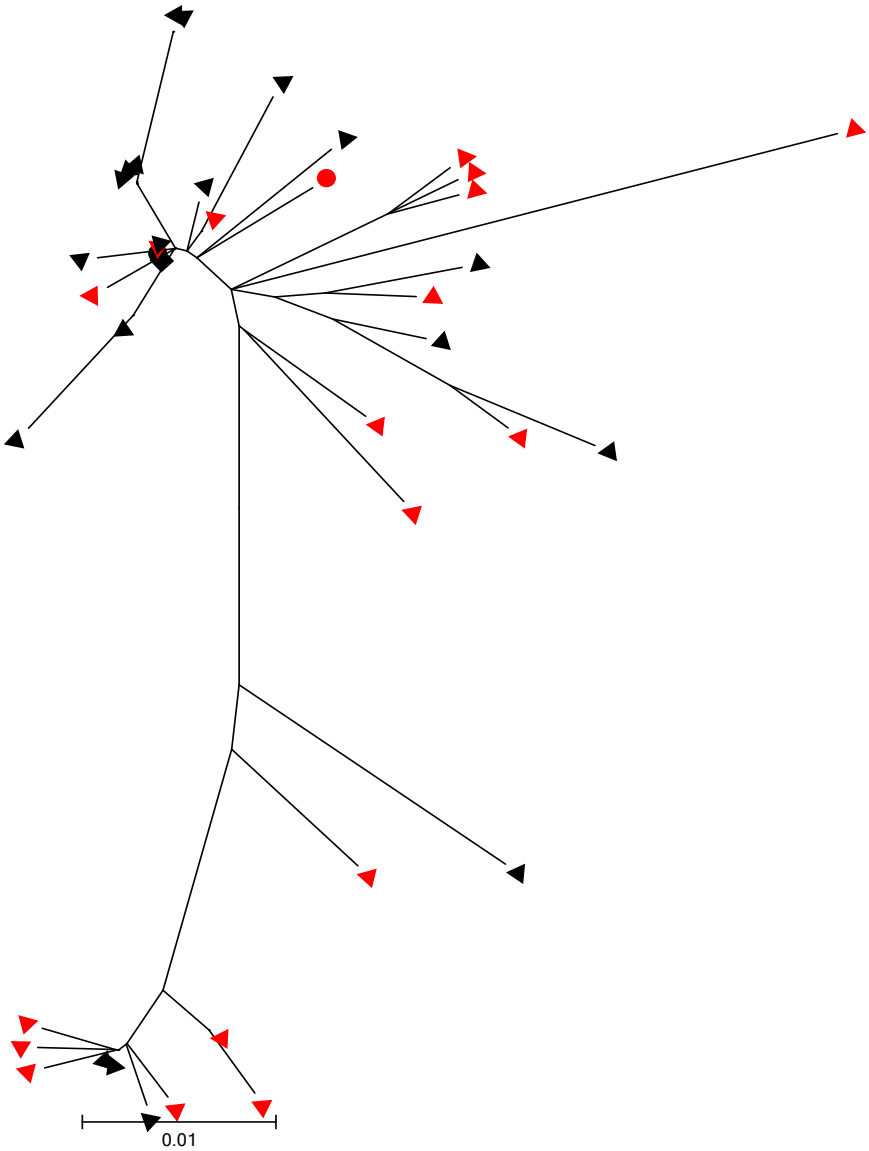


Fig 8.4. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pretreatment sample (red triangles), with the immunoglobulin bound (black square – Week 16/red square – Week 0) and free (black circle – Week 16/red circle – Week 0)sequences included for Subject N. The tree was generated using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5.

Subject N demonstrated immunoglobulin binding of the HVR1 master sequence at week 0, with the IgG free fraction demonstrating a single amino acid difference but by week 16, the master sequence which has persisted throughout the study period appeared to no longer demonstrate IgG affinity (Fig 8.4).

8.3.1.5 Subject T

Fig 8.5. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pretreatment sample (red triangles), with the immunoglobulin bound (black square – Week 16/red square – Week 0) and free (black circle – Week 16/red circle – Week 0) sequences included for Subject T. The tree was generated using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5.

Patient T demonstrated no IgG binding to the sequences produced in the clonal samples but instead binds to a distant genotype 3a sequence (Fig 5).

8.3.2 Subjects with HVR1 change

We included four subjects where the cloning data suggested significant changes in the HVR1 QS but the samples produced following passage through the immunoglobulin depletion kits failed to produce any amplification product in three of these subjects (G, L and Q). Therefore it was only possible to evaluate temporal IgG binding in a single subject where the cloning study had identified HVR1 QS change (subject H).

8.3.2.1 Subject H

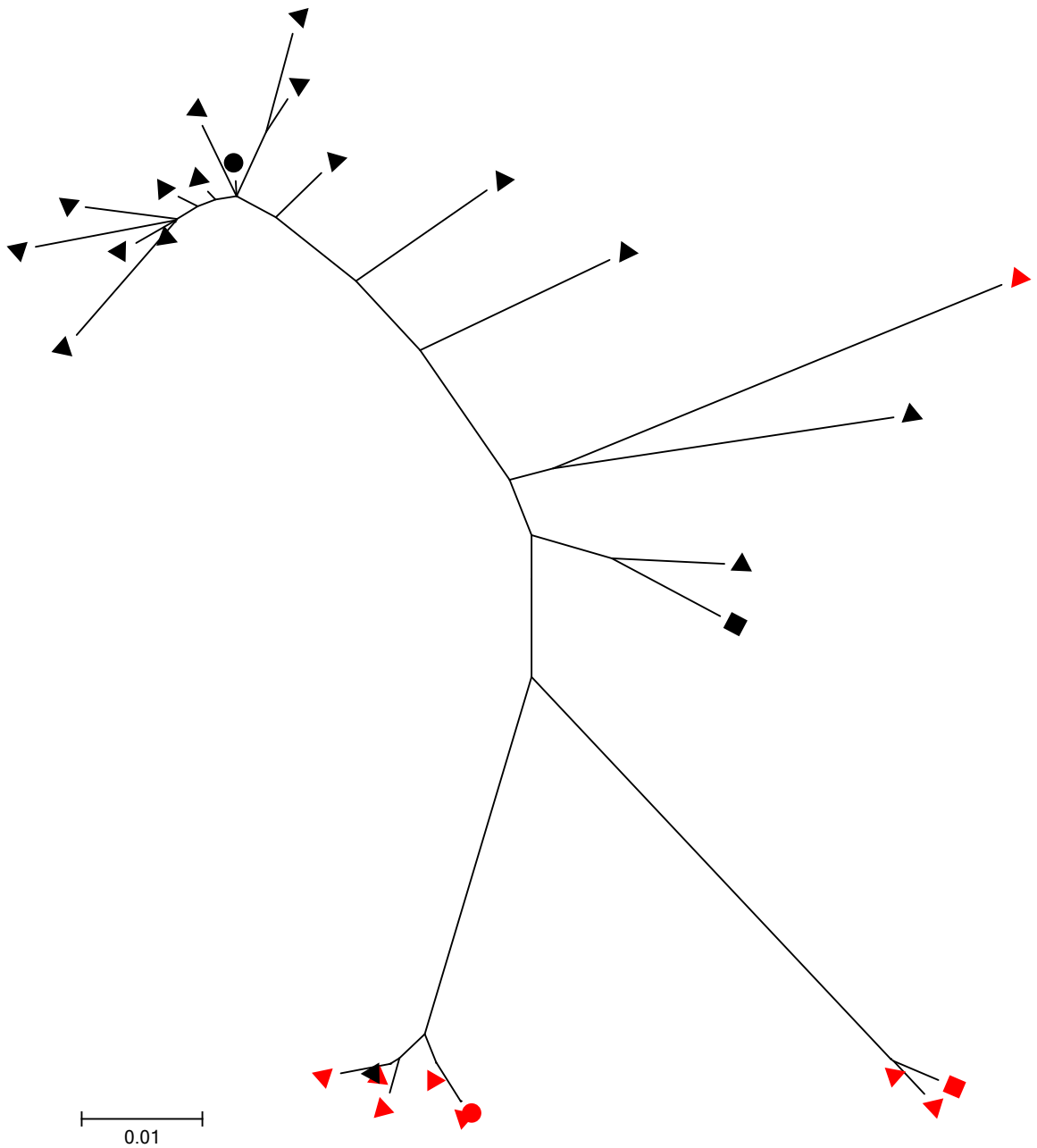


Fig 8.6. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pretreatment sample (red triangles), with the immunoglobulin bound (black square – Week 16/red square – Week 0) and free (black circle – Week 16/red circle – Week 0) sequences included for Subject H. The tree was generated

using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5.

Subject H demonstrates IgG free fractions which correlate with the master sequence at each time point but also IgG bound fractions which correspond with minor sequences from the cloning data (Fig. 6).

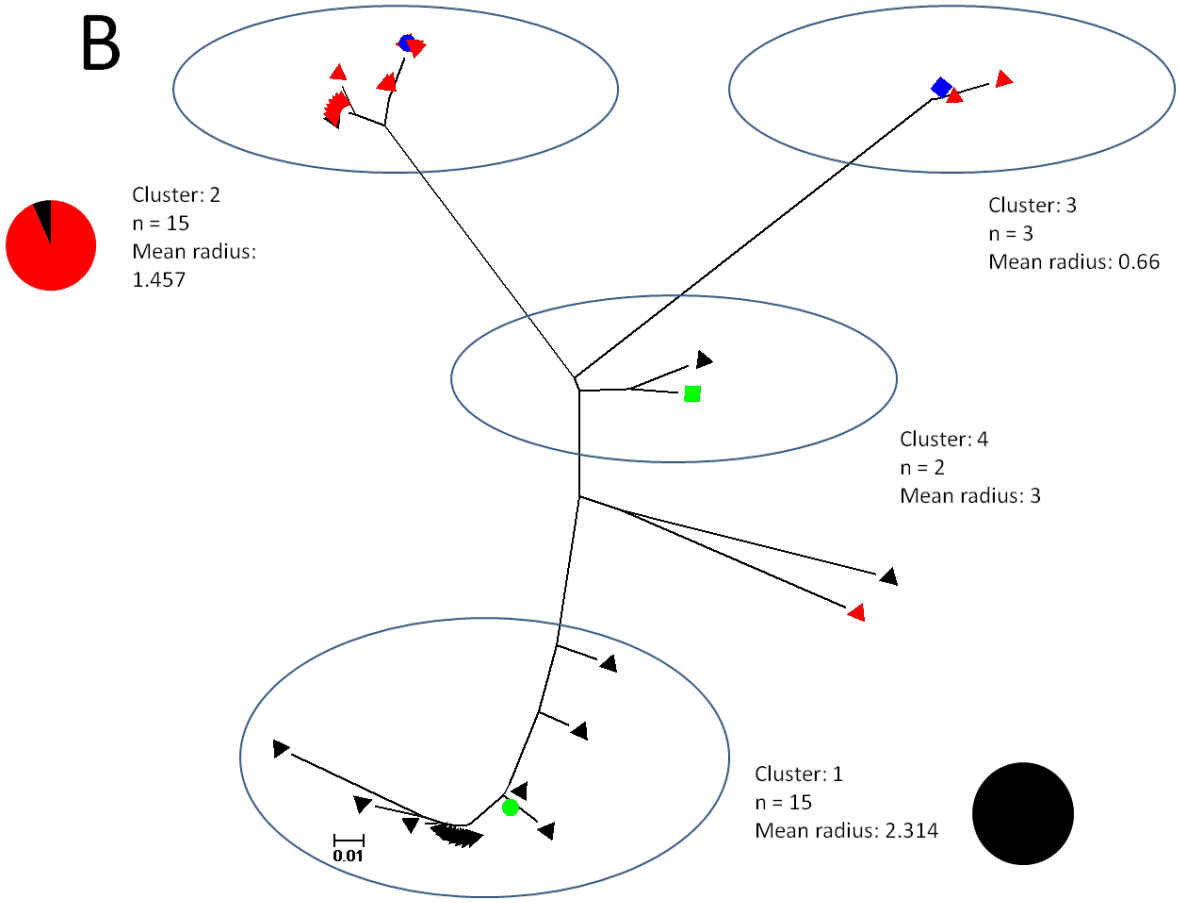


Fig 8.7. Phylogentic tree containing all unique cloned HVR1 nucleotide sequences for samples taken 16 weeks prior to commencing treatment (black triangles) and the pre treatment sample (red triangles). Week 16 IgG bound and free sequences are included and designated with green square and circle respectively. Week 0 IgG bound and free sequences are included and designated with blue square and circle respectively. The tree was generated using maximum composite likelihood (GTR+I+G) with 10,000 bootstrap replicates for tree optimisation using MEGA 5. The subpopulations identified using partitioned analysis are circled and designated 1-4 with the mean radius indicating the mean nucleotide substitutions between all sequences contained within the cluster. Cluster1 and 2 represent the dominant sequences and Venn diagrams are provided indicating the relative proportion of sequences from each sample that make up the cluster.

In order to further characterise the interplay between the sequences generated for subject H, we performed a partitioned analysis of the quasispecies generated. Partitioned analysis of subject B indicates that the HVR1 milieu can be divided into four subpopulations of quasispecies (Fig. 8.7). There is temporal variation in the prevalence of each swarm, with the disappearance IgG enriched cluster from week 16 by the end of the study period and transfer of IgG binding to an entirely new cluster 8 amino acids removed from the original target (Fig. 8.8). This new antibody target is not the dominant QS cluster at week 0 suggesting that it is likely to represent a cluster that became transiently dominant in the intervening period but, by the completion of the study a mature antibody mediated response was in the process of neutralising this cluster. The dominant cluster at the commencement (cluster 1) of the study for which there was no apparent IgG binding has been completely removed from the circulating quasispecies, potentially indicating that a further antibody mediated selective sweep has also neutralised this cluster at some point during the study period (Fig. 8.7).

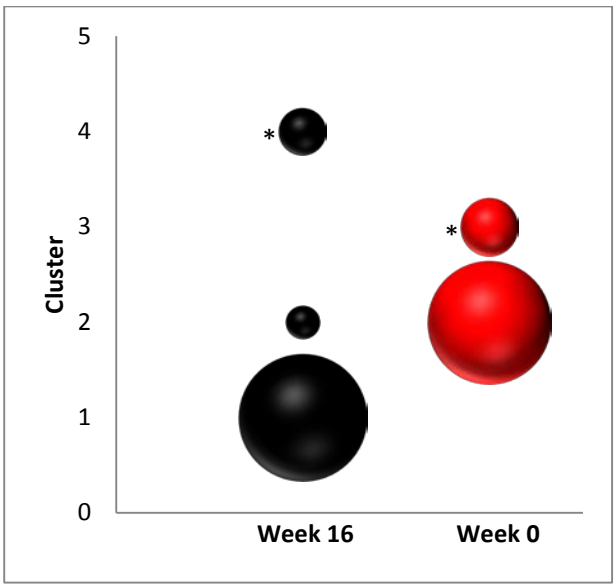


Fig. 8.8 The prevalence of each subpopulation/cluster as identified in Fig 8.7 within the population of HVR1 clones generated during the 16 weeks studied. * designates the clustered targeted by immunoglobulin binding.

Finally, we looked at the prevalence of each cluster over the course of the study (Fig. 8). Cluster 1 the initially dominant cluster is no longer present by week 16 and has been replaced by cluster 2 as the dominant QS. It is notable that neither dominant cluster demonstrates IgG enrichment while both IgG enriched fractions represent minority subsets of the QS in each sample.

8.4 Discussion

HVR1 demonstrates variable patterns of evolutionary change among subjects with chronic infection, with evidence of time order phylogeny over time intervals as short as 2-4 weeks in non cirrhotic patients, and stasis in a proportion of non cirrhotic patients and all cirrhotic patients(302). In order to investigate whether the changes described are due to antibody mediated immune clearance we separated the samples into IgG enriched and IgG depleted fractions in subjects where we had described differing patterns of HVR1 change.

Among subjects where HVR1 stasis was identified among the cloned sequences, variable patterns of IgG binding were described. Subjects A and B demonstrate IgG binding to the master sequence but, no resultant change in the HVR1 quasispecies. Subject N demonstrates a similar lack of interval change in HVR1 sequences but with interval loss of IgG binding to the master sequence. These features are suggestive either of non neutralising antibody response, or binding to a site remote to HVR1. IgG binding in Subject T isolates a remote HVR1 sequence which had not been identified in the cloning study. Interestingly, we have seen how the separation of HCV in infected sera into immunoglobulin bound and separated fractions produce antibody bound sequences which are genetically distantly removed from those generated using clonal analysis. This may represent the persistence at low frequency of previously targeted sequence motifs and may suggest either memory or the ongoing ability of the virus to revert to consensus should pressure to adapt diminish. These low copy viromes may have important implications for resistance to new direct acting anti virals.

The pattern of IgG binding in Subject H, in whom we had identified a time order phylogeny in our cloning data, is highly suggestive of serial emergence of neutralising antibody response with the selection and removal of entire subpopulations from the circulating HVR1 quasispecies milieu driving ongoing sequence divergence. The use of partitioned analysis aids in the identification of this process and highlights the importance of subpopulations in the persistence of chronic infection.

We aimed to study temporal IgG binding to HVR1 in chronically HCV infected individuals with variable patterns of HVR1 change as identified by our prospective cloned sequences. Unfortunately, in three of the four subjects with HVR1 change, the spin columns failed to produce virus particles sufficient for sequencing following reverse transcription and nested PCR. This highlights a potential pitfall in the use of spin columns where the 25µL serum sample may contain insufficient virus particles, particularly if mixing of the serum sample prior to pipetting has been inadequate.

8.5 Conclusion

The separation of HCV quasispecies into antibody enriched and depleted fractions can aid in identifying the predominant evolutionary process driving sequence change. This can assist in identifying suitable subjects for studying virus evolution unaffected by host adaptive immune response. In subjects where host immune response is driving HVR1 change, the bound fraction may point the way of the past and likewise the IgG free fraction may point the way of the future. In the era of direct acting anti virals, our findings raise the prospect that viral resistance may not be easily identifiable in advance of treatment.

Chapter 9

Predicting response to pegylated interferon and ribavirin using HVR1 quasispecies parameters in conjunction with known viral and host markers

9.1. Introduction

Molecular studies investigating response to dual therapy characterised a number of viral and host factors that appeared to suggest that treatment could be individualised at a molecular level. These include

Viral:

1. HVR1 complexity
2. HVR1 diversity
3. Interferon Sensitivity Determining Region
4. HCV Core Sequence

Host:

1. IP 10 levels
2. IL 28 polymorphisms

A comprehensive review of the literature relating to these factors has been included in the thesis introduction (sections 1.7.3 and 1.7.4). We investigated these host and viral molecular markers of treatment response in order to explore possible treatment individualisation strategies.

9.2 Methods

See Section 2.8 methods chapter.

9.3. Results

9.3.1 Core and ISDR Sequences

Subject	Core 70 Sequence Change	Core 91 Sequence Change	ISDR Amino acid Substitutions
A	N	N	1
B	N	N	0
C	N	N	1
D	N	N	0
E	N	N	N/A
F	N	N	0
G	N	N	0
H	N	N	N/A
I	N	N	0
J	N	N	0
K	N	N	N/A
L	N	N	N/A
M	N	N	0
N	N	N	1
O	N	N	0
P	N	N	0
Q	N	N	0
R	N	N	0
S	N	N	0
T	N	N	0
U	N	N	2
V	N	N	0
W	N	N	0

9.3.1 Table 1.

Temporal change in the Core and Interferon Sensitivity Determining Region (ISDR) during the 16 week study period. N signifies no change in the underlying amino acid sequence. N/A signifies no result available. Numbers signify the number of amino acid substitutions that have occurred between the first sample (16 weeks prior to treatment) and the pre treatment sample.

Both Core and ISDR demonstrate far less temporal change when compared with the HVR1 as previously described in Chapters 4-6. There is no change in HCV core sequences at either amino acid position 70 or 91 during the 16 week interval of the study.

There is limited change in the 40 amino acid interferon sensitivity determining region with 3 subjects investigated (A, C, N) demonstrating a single amino acid substitution and one further subject (U) demonstrating two amino acid substitutions during the study period (9.3.1. Table 1.).

Subject	Genotype	Core 70	Core 91
A	1b	Q	C
B	1b	R	M
C	1b	Q	M
D	1b	R	M
E	1b	R	C
F	1b	R	C
G	3a	R*	C *
H	3a	R*	C *
I	3a	R*	C *
J	3a	Q*	C *
K	3a	R*	C *
L	3a	R*	C *
M	3a	R*	C *
N	3a	R*	C *
O	3a	R*	M*
P	3a	Q*	M*
Q	3a	R*	C*
R	3a	R*	C*
S	3a	R*	C*
T	3a	Q*	C*
U	3a	R*	C*
V	3a	R*	C*
W	3a	R*	C*

9.3.1 Table 2.

Amino acid residues at position Core70 and Core91. The amino acids are designated using the IUPAC (International Union of Pure and Applied Chemistry) nomenclature. (C – Cysteine, M – Methionine, R – Arginine, Q – Glutamine). *denotes subjects with genotype 3 infection where no association between specific amino acid substitutions and likelihood of treatment success has been described.

Examination of the Core amino acid sequences identified a number of subjects (highlighted in red) (Table 2) with unfavourable amino acids at positions 70 and or 91 at the commencement of treatment. Three genotype 3 subjects (18%) also had a glutamine at Core 70 one of whom (subject P) also had a methionine substitution at Core 91. Two genotype 3 patients (12%) had a methionine at Core 91. All genotype 3 patients were successfully treated with dual therapy.

Among genotype 1 subjects, 2 (Subjects A and C) (33%) had an unfavourable glutamine substitution at Core 70 with one of these (subject C) also having an unfavourable methionine substitution at Core91. Notably, Subject A achieved SVR but Subject C with both substitutions did not respond to treatment.

Unfavourable methionine residues at Core91 were seen in 3 genotype 1 subjects (50%) and only one of these (Subject D) achieved SVR.

Subject	Amino Acid Substitutions from Wild Type - Week 16	Amino Acid Substitutions from Wild Type - Week 0
A	0	1
B	1	1
C	2	1
D	1	1
E	N/A	N/A
F	1	1
G	6	6
H	N/A	N/A
I	7	7
J	4	4
K	N/A	N/A
L	N/A	N/A
M	8	8
N	6	6
O	7	7
P	11	11
Q	11	11
R	8	8
S	9	9
T	9	9
U	10	10
V	10	10
W	9	9

9.3.1. Table 3

Temporal change in ISDR sequences during 16 weeks pre treatment. The number of amino acid substitutions between the sample for each subjects and the wild type ISDR from HCV-J as described by Enomoto et al.(211) is designated. N/A indicates no data available.

Subject C who went on to have a null response to treatment had an ISDR profile at week 16 prior to treatment commencement that suggests that treatment response was more likely at that time, when compared with the pre treatment sample.

Genotype 3a ISDR equivalent sequences differ from genotype 1b wild type sequences by between 10 and 28% of amino acid residues. This suggests significantly different sequence characteristics which may explain why ISDR is genotype specific.

9.3.2. Temporal Change in IP 10

	IP 10		
	Week 16	Week 8	Week 0
A	887	1813	900
B	1001	2008	1077
C	4757	4551	2812
D	1037	630	764
E	686	748	509
F	2775	1863	2217
G	721	778	271
H	1305	1391	2471
I	1420	796	1495
J	952	1706	1493
K	585	356	473
L	453	1336	1217
M	455	598	355
N	1045		22
O	1768	1117	620
P	761	1901	916
Q	629	1304	1279
R	1702	1625	1705
S	286	281	798
T	679	1301	1167
U	487	488	811
V	1388	1307	1233
W	543	243	348

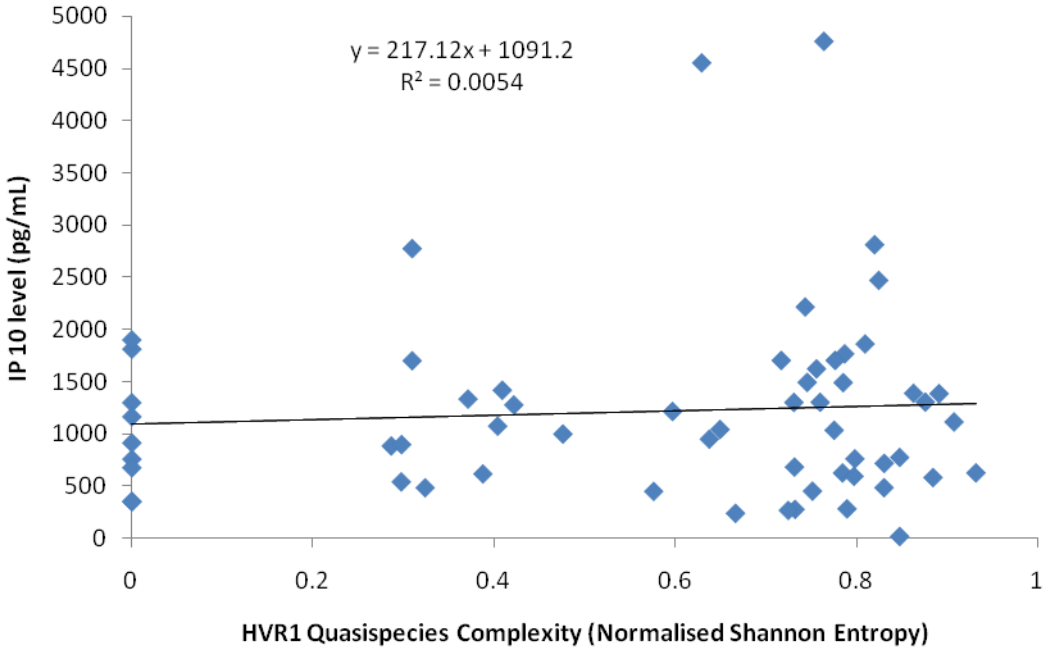
9.3.2. Table 1

Temporal change in Interferon-γ inducible protein 10 kDa (IP-10) for each subject at 16 weeks, 8 weeks and 0 weeks pre-treatment. Samples where the IP 10 level is below the 600pg/mL threshold which has been described as predictive of SVR in genotype 1 patients are highlighted in green.

IP 10 levels behaved in an unpredictable manner with significant changes between samples from the same subject (e.g. Subject I 9.3.2 Table 1). Five subjects transiently had IP 10 levels below the favourable 600pg/mL level described by Lagging et al., compared to only three who demonstrated persistently low IP 10 levels throughout the study suggesting that likelihood of SVR may vary temporally. Among genotype 1 patients, where an association between IP 10 and SVR has been described, only Subject E had a single IP level suggestive of likely treatment success but, this subject did not respond to dual therapy and treatment was discontinued early.

Five of the six patients (83%) with a pre treatment IP 10 level less than 600pg/mL achieved SVR but, this was comparable with the SVR rate for those with an IP 10 level greater than 600pg/mL, where fifteen of 17 patients (88%) achieved SVR. As a result, IP 10 levels below 600pg/mL were not predictive of SVR (χ^2 p=0.8) (9.3.2. Table 1).

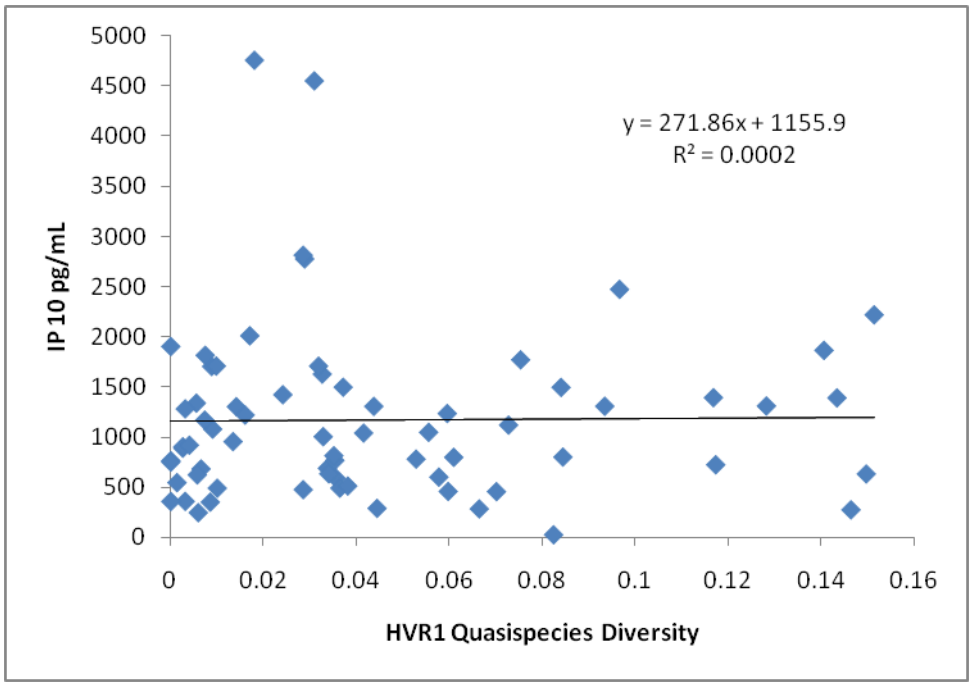
IP 10 levels were lower among non cirrhotic patients (median 934 pg/mL, mean 1055.94pg/mL) when compared with cirrhotic patients (median 1.19pg/mL, mean 1518.17pg/mL) ($p<0.05$) - *Mann Whitney U* difference of two medians.



9.3.2. Fig 1.

Scatter plot with trendline and Pearson correlation of IP 10 levels and HVR1 QS complexity.

Comparison of IP 10 levels with HVR1 QS complexity did not reveal a correlation. In order to clarify whether subjects with very low HVR1 QS complexity (=0) were masking a correlation we also analysed the data excluding these results and it did not alter our finding of no correlation (r=0.073).



9.3.2. Fig 2.

Scatter plot with trendline and Pearson correlation of IP 10 levels and HVR1 QS diversity measured using maximum composite likelihood (GTR+I+G).

HVR1 QS diversity did not correlate with IP 10 levels (r=0.014).

9.3.3. Pre Treatment Prediction Data Summary

Subject	Genotype	Cirrhosis	HVR1 Complexity	HVR1 Diversity	IP-10 <600	IP-10 < 150	ISDR Week -16	ISDR Week 0	Core 70	Core 91	IL-28 rs8099917	IL-28 rs12979860	SVR
A	1b	Y	Y				W	W	Q	C	TG	TC	
B	1b	Y					W	W	R	M	TT	CC	
C	1b	Y					M	W	Q	M	N/A	N/A	
D	1b	Y					W	W	R	M	TG	TC	
E	1b	Y	Y		Y		N/A	N/A	R	C	TG	TC	
F	1b			Y			M	M	R	C	N/A	N/A	
G	3a			Y	Y		M*	M*	R*	C *	TT	TC	
H	3a			Y			N/A	N/A	R*	C *	TG	TC	
I	3a						M*	M*	R*	C *	N/A	N/A	
J	3a			Y			M*	M*	Q*	C *	TT	CC	
K	3a				Y		N/A	N/A	R*	C *	TG	TC	
L	3a						N/A	N/A	R*	C *	TG	TC	
M	3a		Y		Y		M*	M*	R*	C *	N/A	N/A	
N	3a			Y	Y	Y	M*	M*	R*	C *	TT	CC	
O	3a		Y				M*	M*	R*	M*	TG	TC	
P	3a		Y				M*	M*	Q*	M*	N/A	N/A	
Q	3a		Y				M*	M*	R*	C*	N/A	N/A	
R	3a						M*	M*	R*	C*	N/A	N/A	
S	3a			Y			M*	M*	R*	C*	TT	CC	
T	3a	Y	Y				M*	M*	Q*	C*	GG	TT	
U	3a						M*	M*	R*	C*	TT	TC	
V	3a			Y			M*	M*	R*	C*	TT	CC	
W	3a				Y		M*	M*	R*	C*	TT	TC	

9.3.3. Table 1.

Summary of study population including all potential predictors of treatment response and treatment outcome. Characteristics associated with greater likelihood of response are highlighted in green and those associated with less likelihood of response are highlighted in red. *denotes the genotype 3 patients for whom the literature has not previously shown core, ISDR, or IP 10 levels to be predictive of treatment response.

9.3.3. Genotype 1 patients

Five of our six genotype 1 patients were cirrhotic and two cirrhotic patients achieved SVR. Both of these patients had multiple molecular markers that suggested that treatment response was less likely than among their study cohort contemporaries. Conversely, Subject E had the most favourable molecular profile of all cirrhotic genotype 1 patients and, none the less, was a null responder. There were no statistically significant differences between those who achieved SVR and those who did not, though this was most likely the result of insufficient numbers among the cohort.

We evaluated the IL 28 makeup of the host subjects and found that the genotype 1 patient with the most favourable characteristics did not respond to treatment, whereas two of the four subjects with intermediately favourable IL 28 alleles achieved SVR.

9.3.3. Genotype 3 patients

All genotype 3 patients achieved SVR. This precluded analysis of molecular predictors of treatment response.

9.4. Discussion

We investigated host IL28 SNPs, and temporal change in Core70, Core91, ISDR, and Ip 10 levels over a 16 week pre commencement of treatment with dual therapy in a cohort of 23 individuals chronically subjects infected with either genotype 1 or 3, and with differing degrees of liver fibrosis.

Using the data from our cloning study we found no association between either viral or host factors and HVR1 complexity or diversity.

Core and ISDR amino acid substitutions were limited during the study period, though there was a change in the ISDR sequence from favourable to unfavourable in one individual, who subsequently did not respond to treatment, suggesting that the use of this marker requires both a recent sequence result and also, possibly, confirmation on the day of treatment commencement.

The limited degree of change in our Core and ISDR sequences is in keeping with the published literature for these sections of the virome, although no studies evaluating change in either Core or ISDR have to date been published. This suggests that these viral factors are suitable candidates for inclusion in pre treatment prediction models of SVR and their inclusion in broader studies appear likely to be of potential benefit in SVR prediction(412).

We explored the possibility of developing such a model using our data but, were unsuccessful due to the limited numbers in the studied cohort. Of more fundamental importance however in the failure to develop a model was the genotype make-up of the cohort. This was due to both the random presentation of candidates suitable for treatment to our tertiary referral centre and the treatment guidelines at the time the study was undertaken. The study period between 2009 and 2011 corresponded with a period when dual therapy remained the standard of care but the arrival of next generation protease inhibitors, and their improved genotype SVR rates, imminent(413). As a result uncomplicated genotype 1 patients were in many cases awaiting the arrival of protease inhibitors in a process which was named warehousing. This meant that only genotype 1 patients with significant fibrosis/cirrhosis were being treated and this is reflected in the study cohort, with five of the six genotype 1 patients having liver biopsies demonstrating advanced fibrosis(414). The only non cirrhotic genotype 1 patient in our cohort was treated on the request of the patient for occupational reasons.

Conversely, the SVR rates with dual therapy at that time for genotype 3 patients meant that treatment was expedited barring contraindication. This is reflected in the proportion of cirrhotic patients in the genotype 3 study population where only one of seventeen individuals had advanced fibrosis.

Historically, because SVR rates among genotype 3 patients were superior to genotype 1 the discovery of the viral and host characteristics that predict SVR focussed on genotype 1 infection. Consequently,

there are few studies validating these factors in genotype 3 infection. Due to the 100% treatment response in our genotype 3 patients to dual therapy, we were unable to investigate for any associations between these viral and host factors in this study.

We identified an association between increased IP 10 levels and advanced fibrosis but, due to the high proportion of cirrhotic genotype 1 and low proportion of cirrhotic genotype 3 individuals studied we were unable to control for genotype.

9.4.1. Combination studies investigating prediction of treatment response.

A number of studies have been undertaken to investigate whether the viral (Core and ISDR) and host (IP 10 and IL28 SNPs) predictive factors could be combined in order to predict treatment response prior to initiating treatment. In the era of dual therapy, where treatment was associated with significant morbidity, it was argued that this might both minimise patient exposure to unwanted side effects and maximise societal gain through optimal use of resources (in this case the cost of treatment). Accordingly, pre treatment prediction of SVR would aid greatly in resource allocation.

Numerous studies have investigated the utility of combining IL28 SNPs with IP 10 levels to predict treatment response. In most cases, multivariate analysis suggests that IL28 SNPs and IP 10 levels independently predict SVR and that combined they can potentially better predict SVR(393, 415). A number of algorithms have been proposed including these factors with others such as race, age, and/or baseline viral load but low positive (50-80%) and negative (50-70%) predictive values are likely to prevent their widespread use(416, 417).

Studies investigating combined Core and ISDR substitutions suggest that substitutions at Core70 and more than two substitutions in the ISDR are independent predictors of SVR among Japanese and Thai patients on multivariate analysis (418, 419).

The advent of triple treatment with protease inhibitors prompted an initial period of evaluation of these viral and host factors among genotype 1 patients but the swift arrival of new highly efficacious direct acting anti viral medications resulted in a paradigm shift in treatment. This new interferon free era of HCV treatment with pangenotypic SVR rates exceeding 80%, even among patients who would have previously been difficult to treat, and with the added benefit of fewer side effects has threatened to make interferon obsolete in the management of HCV. One major concern with the new DAAs has been the cost which has threatened to fundamentally undermine the health budgets of developed countries and is likely to preclude access to these medications in many developing countries, where the prevalence of HCV is highest(420). In light of these cost concerns, the pre treatment identification

of patients for whom dual therapy is likely to achieve comparable SVR rates, at a fraction of the cost, remains desirable.

9.5. Conclusion

We have identified little temporal change in amino acid sequences in the Core and ISDR regions during the 16 week pre treatment period studied. Although this stability suggests that these viral predictors of SVR could be of significant use in developing models to optimise treatment outcomes, we were unable to demonstrate this in our cohort.

This contrasts with HVR1 change where the time required to generate these metrics is likely to render the finding obsolete.

We found no association between Core, ISDR, IP 10, or IL28 SNPs and either diversity or complexity suggesting that these cannot be used as surrogate markers.

The advent of highly efficacious DAAs is likely to obviate the requirement of pre treatment predictive models on the basis of these historical molecular markers of response to dual therapy, though limited access to these expensive medications in developing economies may ensure a preserved role for interferon. Nevertheless, the potential for HCV adaptation to these drugs and the emergence of DAA resistant mutants remains a real concern and may lead to the re introduction of interferon in combination with DAAs in the future.

Chapter 10

10.1 Conclusion

Using Sanger sequencing of both the nested PCR amplicon and amplified plasmid clones, and next generation sequencing of samples collected over 16 weeks, and with the inclusion of further sequenced plasmid clone sequences from a retrospective sample, we have described HVR1 evolution in unique detail identifying novel patterns of sequence change. Efforts to develop pre treatment prediction models of treatment success were hindered by the genotype make up of the study cohort, itself dependent on the presentation of candidates suitable for treatment. Temporal stasis in core sequences and limited ISDR change suggest that these are suitable candidates for such prediction models. IP 10 levels although statistically lower among cirrhotic patients, were subject to significant unpredictable temporal change which may preclude it from use in pre treatment prediction.

Using numerous modelling strategies including phylogenetics (using evolutionary models informed by jmodelttest), median joining networks , k and one step network techniques, partitioned analysis of quasispecies, and Bayesian techniques to identify nucleotide substitution rates we have developed a robust and accurate schema for analysing and describing patterns and mechanisms of both short and long interval HVR1 evolution and adaptation.

Early chronic HCV infection HVR1 evolution is characterised by multi-lineage episodic divergent evolution characterised by positive selection which is likely to be convergent to consensus at sites not targeted by the adaptive immune response. With prolonged chronic infection HVR1 transitions to single lineage sequence infection with sequence stability which is associated, though not exclusively seen, with advanced liver disease and cirrhosis.

These features are highly suggestive of viral discovery and exploitation of niche deficits in the host immune response.

Cloning depth of 10-20 sequences was sufficient to correctly identify the patterns of HVR1 change seen using next generation sequencing in all but one subject of the 15 studied. The prolonged HVR1 sequence inertia seen in the subject infected with contaminated anti-D immunoglobulin suggests that original antigenic sin combined with the exhaustion of host adaptive immune response may conspire to flatten the fitness landscape and facilitate the exploration of broadening fitness peaks as they transition to plateaus.

IgG fractionation of the HVR1 quasispecies highlights the complexity of HCV antibody mediated clearance. Antigenic drift attributable to adaptive humoral response contrasts with the absence of discernable antibody binding in HVR1 stasis. However, not all rapidly changing HVR1 haplotypes demonstrate antibody binding suggesting that alternative mechanisms of HVR1 change potentially including the emergence of fitter variants or local T-cell mediated clearance of a HVR1 profile compartmentalised within a section of the liver may be responsible. Conversely, antibody binding is not universally associated the exclusion of the HVR1 motif from the quasispecies implying a non neutralising or slowly neutralising antibody response.

Our data illustrates the heterogeneous tempo of nucleotide substitution rates between subjects which is independent of the presence of multiple lineages, presence of advanced fibrosis or time order phylogenetic change. Interestingly low substitution rates are not exclusively the remit of cirrhotic patients but do correspond with subjects where HVR1 change during the study period is limited implying a well adapted haplotype to host environmental truncation of the fitness landscape.

The observation of three apparent mean modes of distribution of nucleotide substitution rate which correlate between HVR1 and E1 (which is not under demonstrable selective pressure in any subject) may be suggestive of adaptation and optimisation of the RNA dependent RNA polymerase mutation rate as we had hypothesized.

New direct acting anti virals have changed the paradigm of expected treatment outcomes in HCV but the emergence of treatment resistant polymorphisms remains a significant concern. Our HVR1 data suggests limitations associated with amplicon sequencing when trying to describe the underlying quasispecies milieu. In screening for resistant polymorphisms we face the dilemma of limitations associated with amplicon sequencing strategies and the economic and time intensive constraints associated with both cloning and next generation sequencing.

Nevertheless, our next generation sequencing network analysis highlights lineages connected by a maximum of one amino acid substitution to the master sequence even among multiple lineage infections. This suggests that the underlying presence of genetically remote low copy memory/resistant genomes capable of expansion to undermine treatment is unlikely. Therefore, virus memory and so called convergent evolution reflect short interval adaptation to global fitness optima rather than an illustration of virus memory.

Furthermore, the emergence of on treatment resistant mutants is likely to result in fitness cost with high probability of reversion to wild type once this selective pressure is removed suggesting that re treatment may not automatically fail. Unlike retroviruses such as HIV where resistance fixation and

transmission is common, the tendency for HCV to converge to consensus is likely to abrogate emergence of resistance to direct acting anti virals on a global scale among treatment naive patients.

Appendix A

Genbank Accession Numbers

Subject A

Retro	KU897268-KU897282
16	KF133681-KF133662
14	KF133703-KF133682
10	KF133725-KF133704
8	KF133743-KF133726
4	KF133763-KF133744
2	KF133787-KF133764
0	KF133806-KF133788

Subject B

Retro	KU897226- KU897244
16	HQ661513
	HQ661501-HQ661508
	HQ661477-HQ661484
	Amplicon - HQ661470
14	KU897245- KU897251
	KU897266- KU897267
12	HQ661488-HQ661492
	Amplicon - HQ661475
6	HQ661509-HQ661512
	HQ661476
	Amplicon - HQ661472
4	HQ661485-HQ661487
	HQ661493-HQ661500
	Amplicon-HQ661473
0	KU897252- KU897265

Subject C

OB	KU897095-KU897111

OA	KU897118-KU897140
O	KU897112-KU897117
O-16	KU897155-KU897171
O-14	KU897172-KU897186
O-12	KU897187- KU897206
O-8	KU897207- KU897225
O-0	KU897141- KU897154

Subject D

Retro	KU897404- KU897433
16	KU897383- KU897394
14	KU897300- KU897316
12	KU897283
	KU897317- KU897329
	KU897402- KU897403
8	KU897330- KU897334
	KU897379- KU897382
	KU897395- KU897401
6	KU897335- KU897342
	KU897372- KU897378
4	KU897343- KU897344
	KU897360- KU897371
2	KU897284- KU897299
0	KU897345- KU897359

Subject E

16	KU897442- KU897458
14	KU897481- KU897502
12	KU897434- KU897441
10	KU897503- KU897519
6	KU897520- KU897534

4	KU897535- KU897550
2	KU897459- KU897480

Subject F

Retro	KU897551-KU897570
16	HQ661524-HQ661541
	Amplicon - HQ661518
14	KU897600
	KU897629- KU897651
12	KU897601
	KU897615- KU897628
	Amplicon – KU661517
8	HQ661541-HQ661547
	HQ661520-HQ661523
	Amplicon - HQ661516
6	KU897602- KU897614
4	KU897591- KU897599
	Amplicon – KU661514
2	KU897571- KU897590
0	HQ661548-HQ661563
	HQ661519
	Amplicon - HQ661515

Subject G

Retro	KU897652- KU897670
16	KC967602-KC967621
14	KC967684-KC967601
10	KC967562-KC967583
8	KC967543-KC967561
6	KC967527-KC967542

4	KC967512-KC967526
2	KC967489-KC967511
0	KC967482-KC967488

Subject H

Retro	KU897671- KU897686
16	KC997241-KC997257
14	KC997258-KC997274
12	KC997275-KC997286
10	KC997288-KC997307
8	KC997308-KC997321
6	KC997322-KC997339
4	KC997340-KC997370
2	KC997371-KC997391
0	KC997392-KC997408

Subject I

Retro	KU897687- KU897698
16	KC997409-KC997429
14	KC997430-KC997450
12	KC997451-KC997467
10	KC997468-KC997489
6	KC997490-KC997512
4	KC997513-KC997531
2	KC997532-KC997552
0	KC997553-KC997568

Subject J

Retro	KU897699- KU897721
16	KF133357-KF133373
14	KF133374-KF133391
12	KF1333920KF133400
10	KF133401-KF133422
8	KF133423-KF133437
6	KF133438-KF133456
4	KF133457-KF133476
2	KF133478-KF133494
0	KF133495-KF133512

Subject K

Retro	KU897723- KU897737
14	KU897722
KU	897738- KU897753
12	KU897754- KU897762
	KU897819
	KU897836- KU897849
10	KU897763- KU897780
8	KU897820- KU897835
6	KU897781- KU897782
	KU897808- KU897818
4	KU897783- KU897788
	KU897795- KU897807
2	KU897789- KU897794

Subject L

16	HQ661684-HQ661701
	HQ661646
	Amplicon - HQ661642

14	KU897850- KU897861
10	KU897862- KU897866
	KU897907- KU897916
8	HQ661652-HQ661659
	Amplicon - HQ661643
6	KU897867- KU897872
	KU897892- KU897906
4	HQ661660-HQ661665
	HQ661647-HQ661651
	Amplicon - HQ661644
2	KU897873- KU897891
0	HQ661666-HQ661683
	Amplicon - HQ661645

Subject M

Retro	KU897949-KU897968
16	HQ661599-HQ661615
	Amplicon - HQ661569
12	HQ661631-HQ661641
	Amplicon - HQ661568
10	KU897918
	KU897938- KU897948
	KU897969-KU897971
8	HQ661616-HQ661630
	Amplicon - HQ661567
6	KU897919- KU897937
4	HQ661584-HQ661598
	Amplicon - HQ661564
2	KU897917
	KU897972-KU897989

	Amplicon - HQ661566
0	HQ661599-HQ661615
	Amplicon - HQ661565

Subject N

Retro	KU897991- KU898003
16	HQ661715-HQ661734
	Amplicon - HQ661702
2	KU897990
	KU898004-KU898022
	Amplicon - HQ661703
0	HQ661735-HQ661743
	HQ661705-HQ661714
	Amplicon-HQ661704

Subject O

Retro	KU898023-KU898039
16	KF133513-KF133529
14	KF133530-KF133550
12	KF133551-KF133557
10	KF133558-KF133581
8	KF133582-KF133599
6	KF133600-KF133613
4	KF133614-KF133628
2	KF133629-KF133644
0	KF133645-KF133661

Subject P

Retro	KU957001-KU957019
16	KU956994

	KU957026-KU957029
	KU957146-KU957155
12	KU956995-KU957000
	KU957020-KU957025
10	KU957030-KU957034
	KU957131-KU957145
8	KU957035-KU957046
	KU957068
	KU957125-KU957130
6	KU957047-KU957067
4	KU957069-KU957081
	KU957122-KU957124
2	KU957082-KU957088
	KU957107-KU957121
0	KU957089-KU957106

Subject Q

Retro	KU957156-KU957169
16	HQ661460-HQ661469
14	KU957171-KU957190
12	HQ661447-HQ661459
10	KU957191-KU957196
8	HQ661428-HQ661446
6	KU957170
4	HQ661410-HQ661427
2	KU957197-KU957211
0	HQ661391-HQ661409

Subject R

Retro	KU957669-KU957684

16	KU957667
	KU957755-KU957771
14	KU957668
	KU957685-KU957686
	KU957735-KU957754
12	KU957650-KU957666
10	KU957687-KU957690
	KU957721-KU957734
8	KU957691-KU957694
	KU957709-KU957720
6	KU957695-KU957708
4	KU957642-KU957649
	KU957772-KU957796
0	KU957639-KU957641
	KU957797-KU957810

Subject S

16	KU957213-KU957218
14	KU957212
	KU957219
	KU957234
	KU957254
	KU957303-KU957311
12	KU957271-KU957275
	KU957290-KU957302
8	KU957276-KU957289
6	KU957220-KU957233
4	KU957235-KU957253
2	KU957255-KU957270

Subject T

16	HQ661756-HQ661770
	Amplicon - HQ661744
14	KU957602-KU957615
12	KU957554-KU957569
	Amplicon – HQ661745
8	HQ661780-HQ661791
	HQ661749-HQ661749
	Amplicon - HQ661746
6	KU957616-KU957638
4	KU957539-KU957553
	KU957570-KU957589
	Amplicon – HQ661747
2	KU957525-KU957538
	KU957590-KU957601
0	HQ661771-HQ661779
	HQ661750-HQ661755
	Amplicon - HQ661748

Subject U

Retro	KU957312-KU957327
16	KC964872-KC964893
14	KC964894-KC964913
12	KC964914-KC964937
10	KC964938-KC964959
8	KC964960-KC964981
6	KC964982-KC965000
4	KC965002-KC965024
2	KC965025-KC965046

Subject V

Retro	KU957328-KU957344
14	KC953964-KC953984
12	KC951985-KC954007
10	KC965008-KC954029
8	KC954030-KC954051
6	KC954052-KC954068
2	KC954069-KC954091

Subject W

Retro	KU957509-KU957524
16	KU957366-KU957384
14	KU957345-KU957365
12	KU957385-KU957397
10	KU957398-KU957418
8	KU957419-KU957441
4	KU957442-KU957464
2	KU957465-KU957484
0	KU957485-KU957508

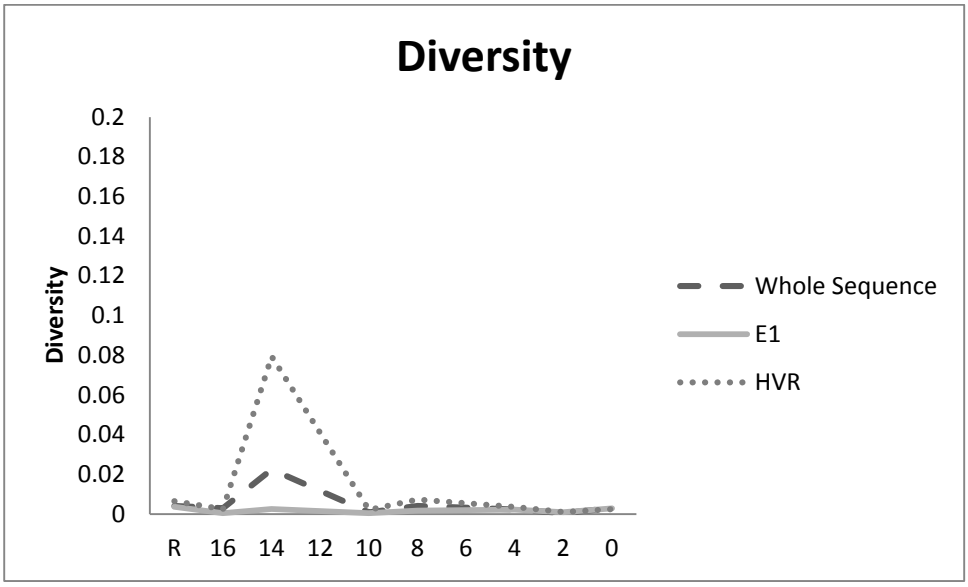
Pyrosequencing data sets available at http://www.ucc.ie/liamfanning/hcv

Appendix B

Retrospective data for remaining Subjects

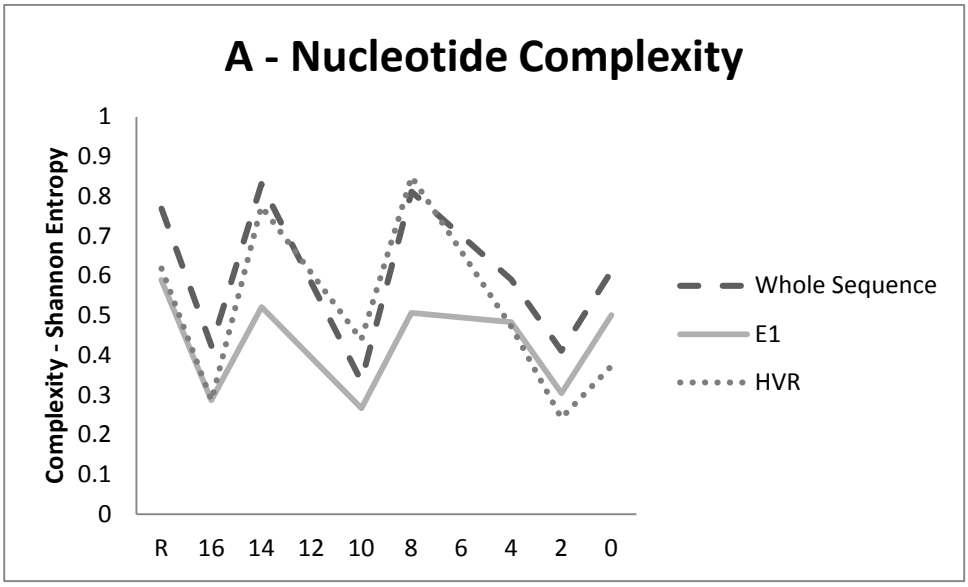
A.1. Subject A

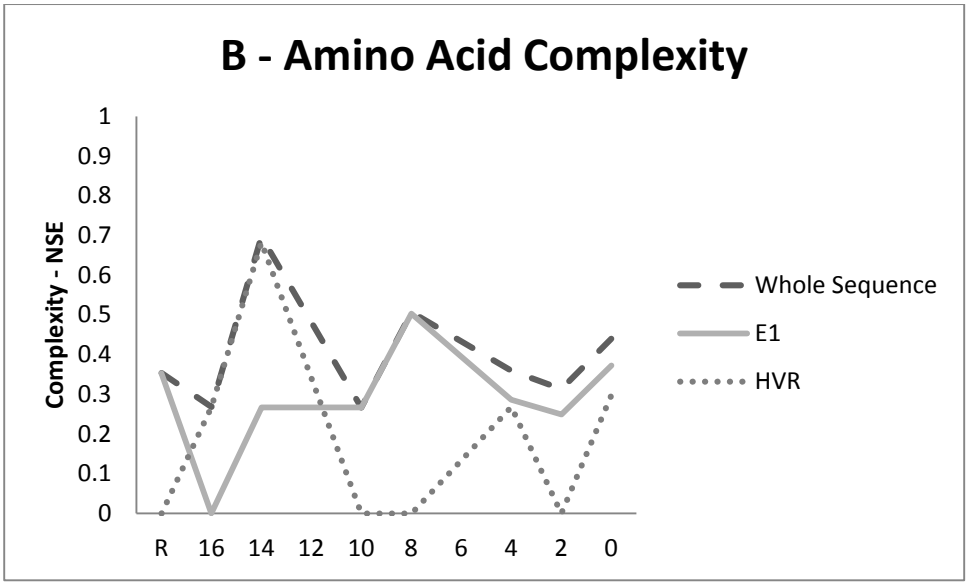
A.1 Diversity, Complexity, and Divergence



A.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

HVR1 diversity is greater than E1 diversity but to a far less degree than in other subjects with the exception of the sample corresponding with week 14 where there is marked HVR1 diversity.

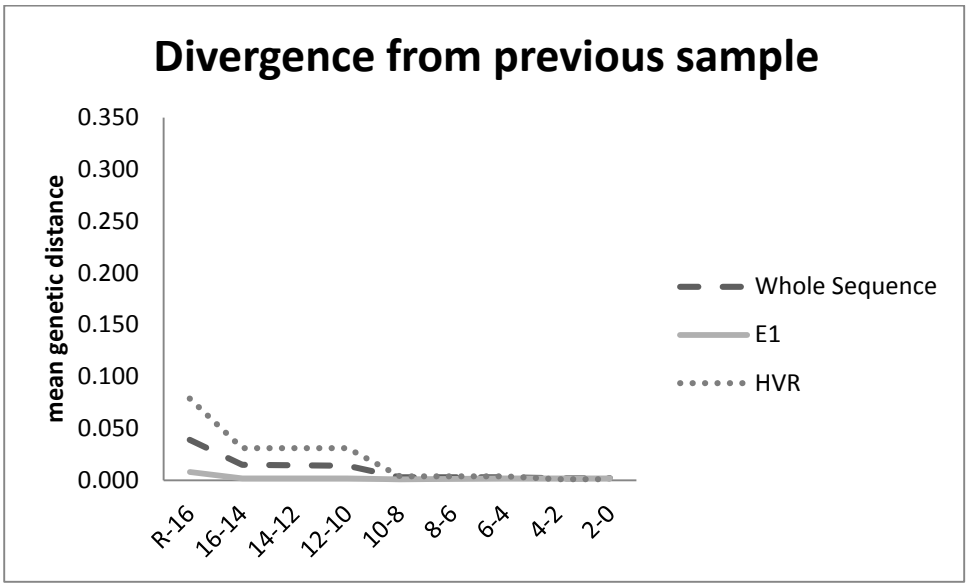




A.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

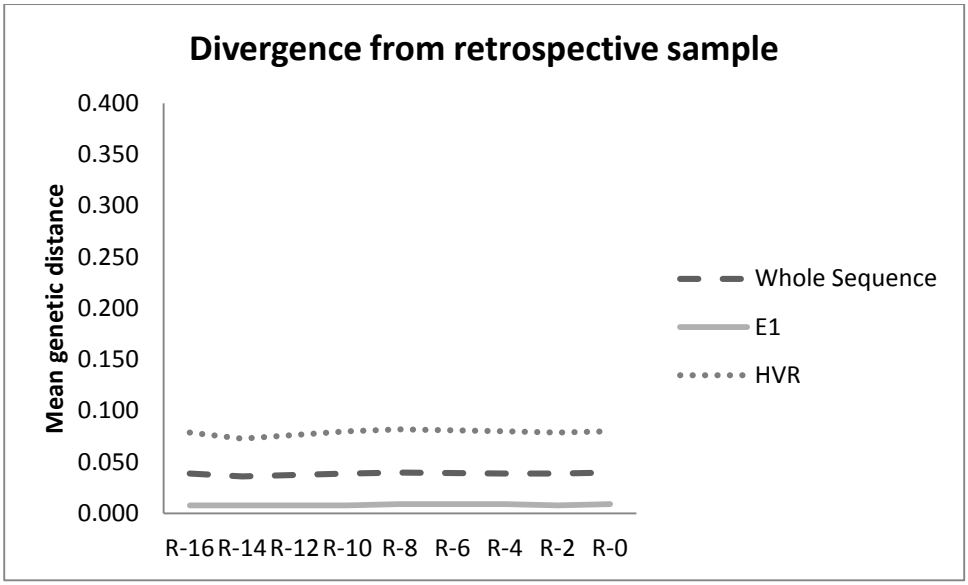
HVR1 demonstrates complexity which is similar in most samples to E1 which is unusual.

Additionally, HVR1 amino acid complexity is less than E1 in most samples which is not seen in any other subject.



A.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

Divergence between samples is maximal between the retrospective sample and the week 16 sample with little subsequent divergence in HVR1 through the remainder of the study.



A.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective is almost equal for each subsequent sample suggesting QS stasis.

A.1 Phylogenetic analysis

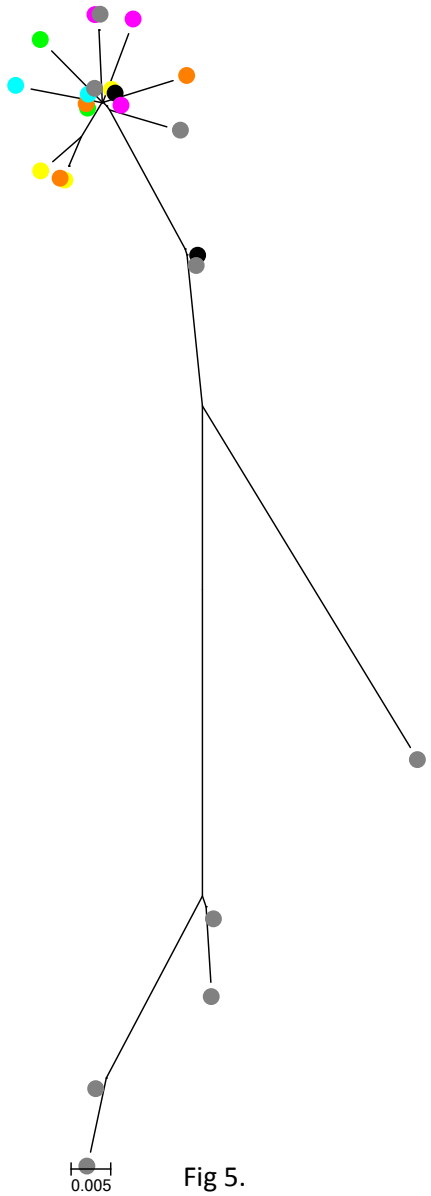


Fig 5.

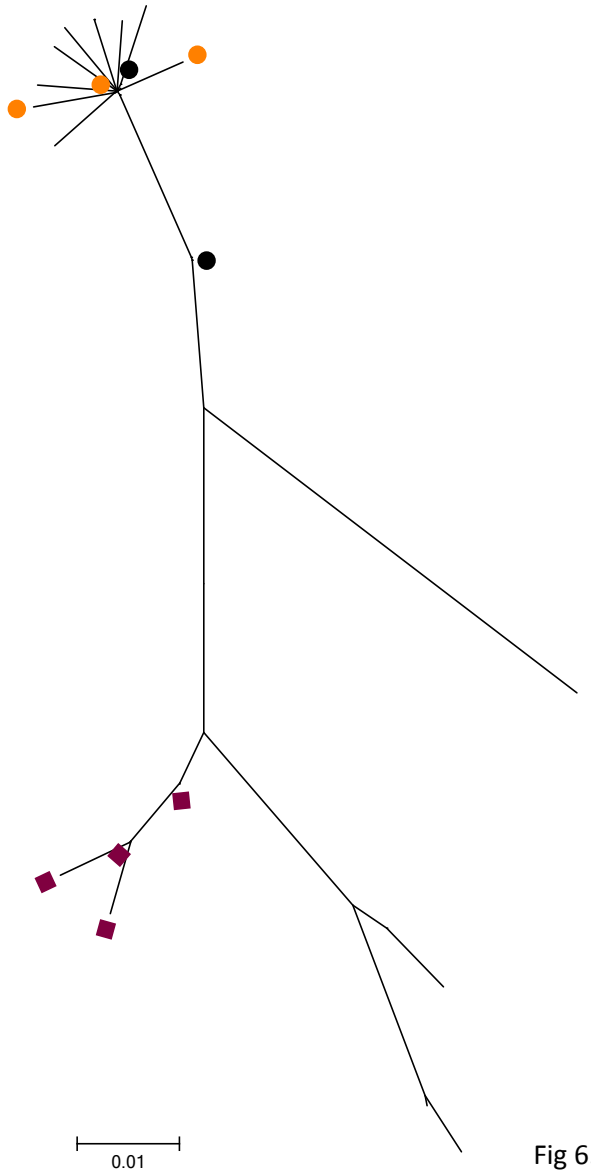


Fig 6.

A.1 Fig 5. Phylogenetic tree produced when all unique HVR1 sequences were included for the 16 weeks prior to commencing treatment. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 10 – green, Week 8 – yellow, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.

A.1. Fig 6. Phylogenetic tree with all unique HVR1 sequences with retrospective (378 days prior to Week 16 sample) and samples from week 16 and week 0 labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation.

It is noticeable that the general shape of the tree is altered by the inclusion of the retrospective

sample with the emergence of a new clade.A.1 Subpopulation analysis

```
                        1        10         20      27
                        |--------+---------+------|
1212351_M13uni_--_29    QTYATGGTQAHTTNRLVSFLSPGPSQK
J-2-10_M13uni_--_18.    QTYATGGTQAHTTNRLVSFLSPGPSQK
12123513_M13uni_--_1    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-4-1_M13uni_--_1    QTYATGGTQAHTTNRLVSFLSPGPSQK
J-2-1_M13uni_--_10..    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-4-11_M13uni_--_    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212161_M13uni_--_29    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-82_M13uni_--_13    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-143_M13uni_--_2    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-84_M13uni_--_14    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-86_M13uni_--_14    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-101_M13uni_--_1    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-1020_M13uni_--_    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-141_M13uni_--_9    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212-1410_M13uni_--_    QTYATGGTQAHTTNRLVSFLSPGPSQK
1212355_M13uni_--_10    QTYATGGTQAHTTNRLVSFLSPGPPQK
1212-4-10_M13uni_--_    QTYATGGTQAHTTNRLVSFLSSGPSQK
1212-1412_M13uni_--_    QTYVTGGTQAYTTNRLVSFLSPGPSQK
12121613_M13uni_--_6    QTYVTGGTQAYTTNRLVSFLSPGPSQK
1212-1415_M13uni_--_    QTRVTGGTQAFITNRLVSFLTPGPSQK
1212-1414_M13uni_--_    RTDVVGGTQAFTTNRLAGFLSPGPSQK
1212-1417_M13uni_--_    RTDVVGGTQAFTTNRLAGFLSPGPSQK
1212-1421_M13uni_--_    RTDVVGGTQAFTTNRLAGFFSPGPSQK
1212-146_M13uni_--_1    QTDVVGGTQAFTTNRLAGFFSPGPSQK
J-10_M13uni-21_--_8.    RTNVMGGTQAYTTNRLVGFLSPGPSQK
J-14_M13uni-21_--_8.    RTNVMGGTQAYTTNRLVGFLSPGPSQK
J-15_M13uni-21_--_8.    RTNVMGGTQAYTTNRLVGFLSPGPSQK
J-24_M13uni-21_--_14    RTNVMGGTQAYTTNRLVGFLSPGPSQK
           Consensus    qT.vtGGTQA.tTNRLvsFlsPGPSQK
```

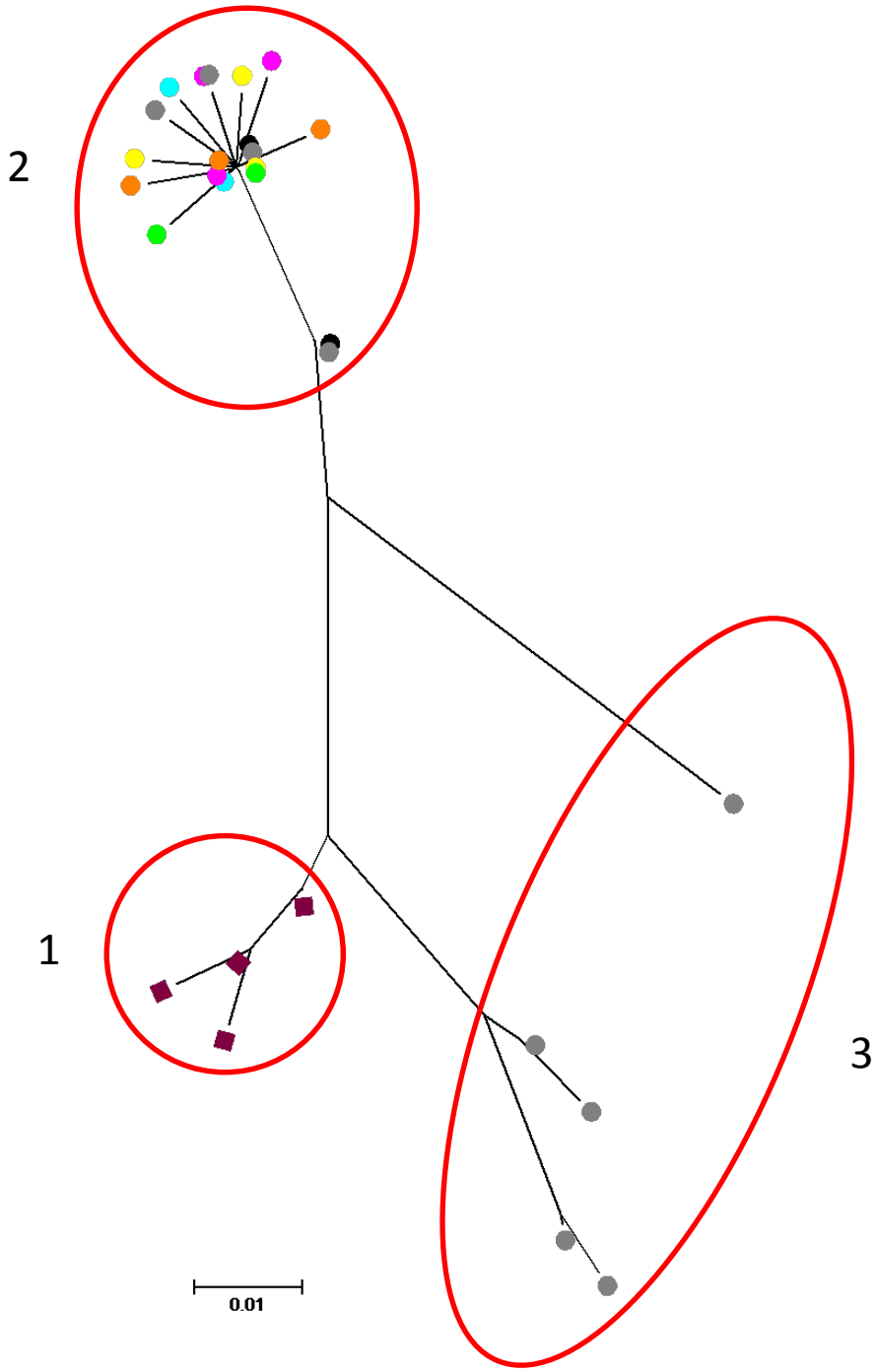
2

3

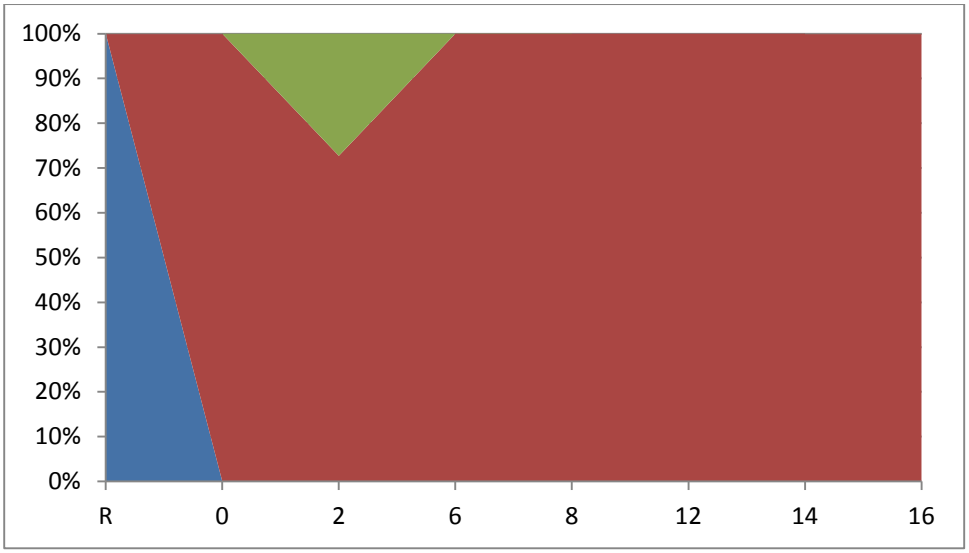
1

A.1. Fig 7 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study.

This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (3 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which had the higher number of sequences was labelled first. This clearly illustrates that the inclusion of the retrospective sample results in a new subpopulation.



A.1. Fig 8 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 10 – green, Week 8 – yellow, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.

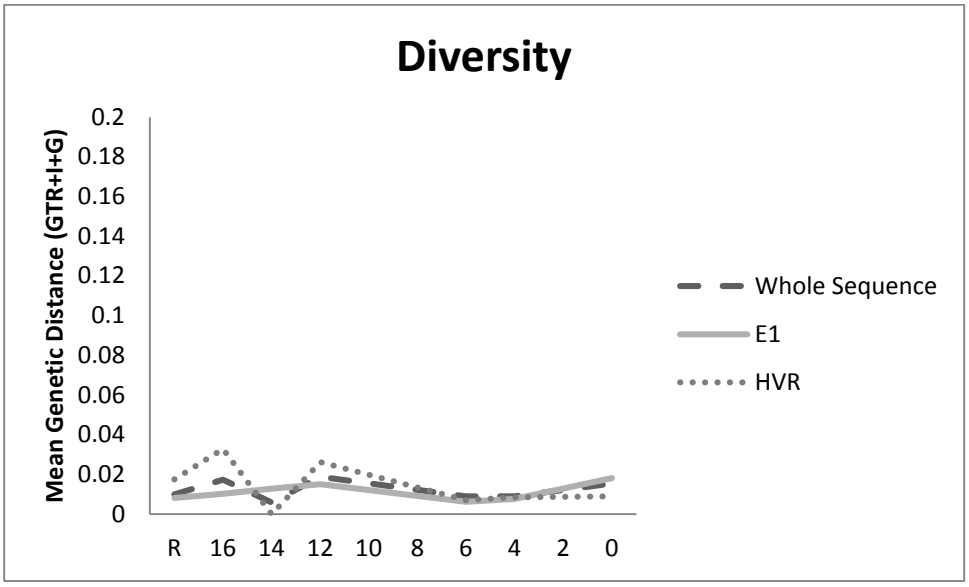


A.1. Fig. 9.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

The initial dominant QS subpopulation (1) is no longer present by the time the first study sample was collected one year later and 16 weeks prior to treatment. Subpopulation 3, seen here in green, transiently appears in a single sample. Examination of the phylogenetic tree indicates that this clade was likely the result of alternate sampling of the sequence space but it appears that this clade has been excluded by superior fitness of subpopulation 2.

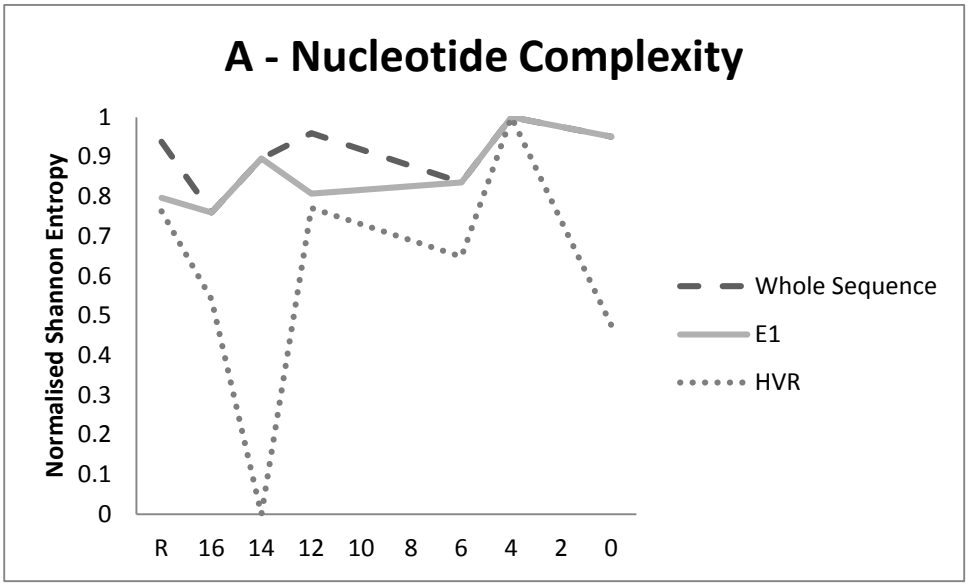
B.1 Subject B

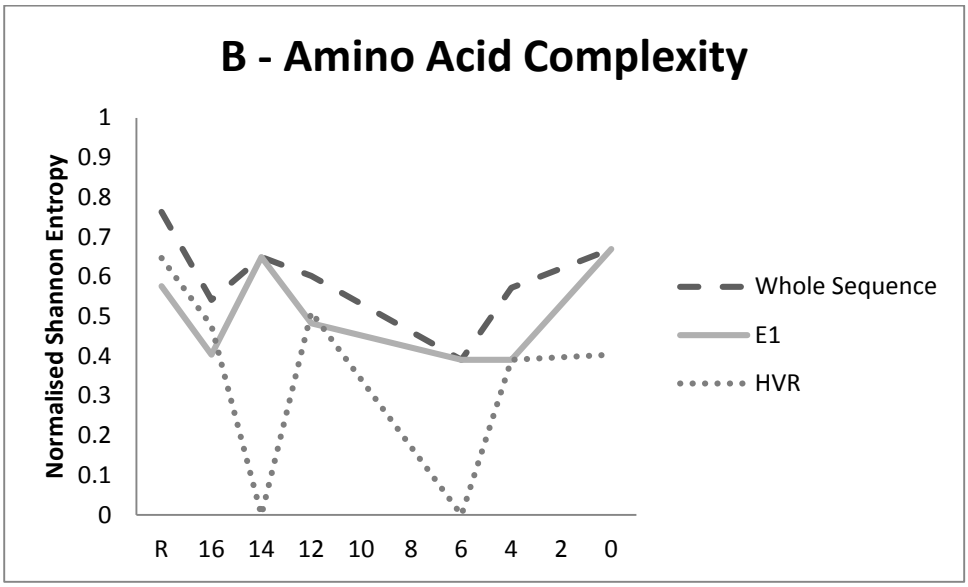
B.1 Diversity, Complexity, and Divergence



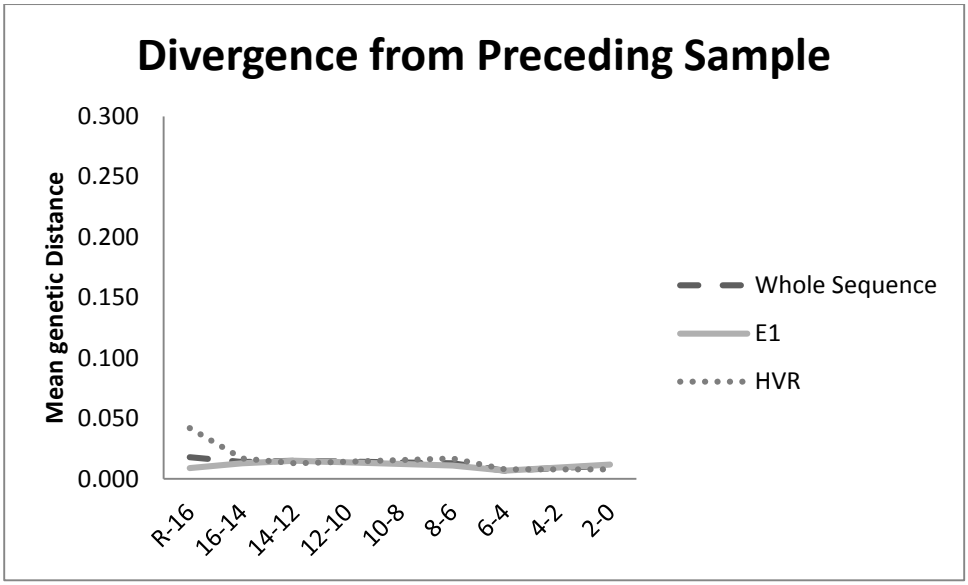
B.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

HVR1 diversity is greater than E1 diversity.



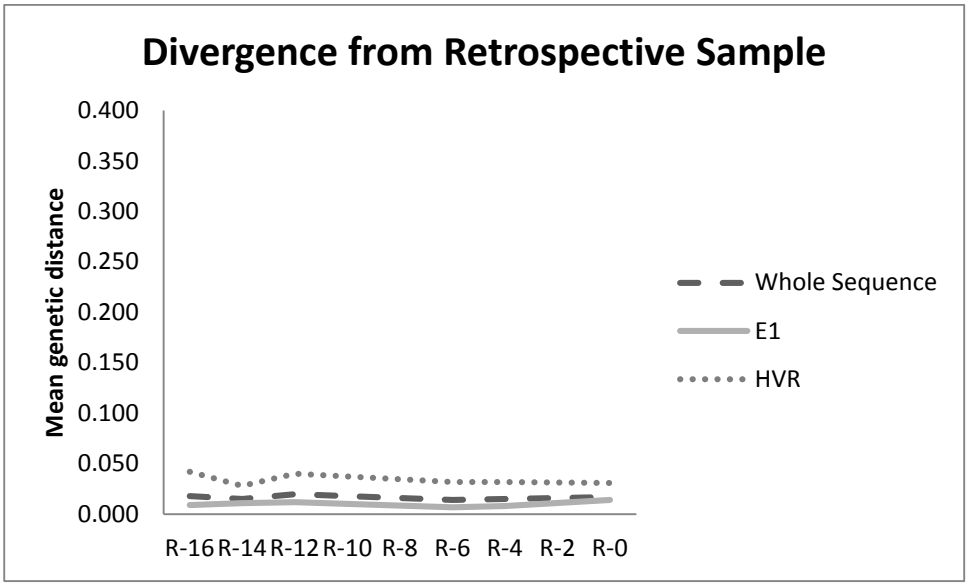


B.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.



B.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

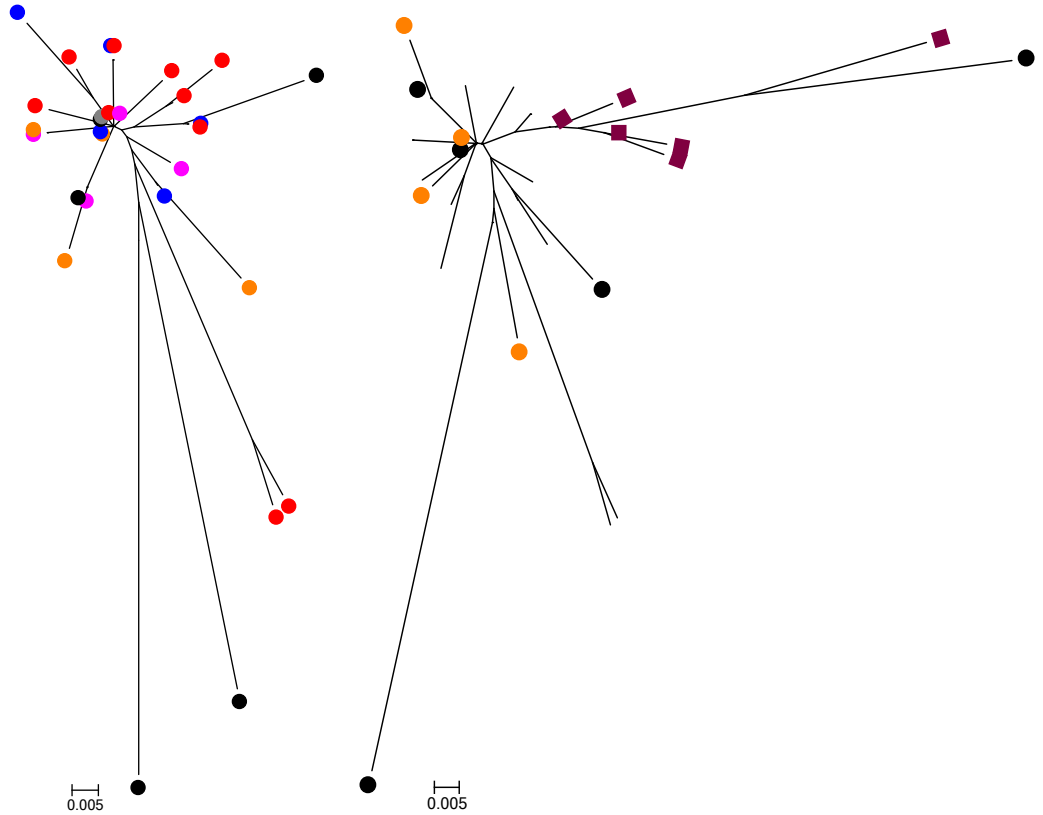
Subject B demonstrates minimal divergence throughout the study, including the retrospective sample.



B.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

Both HVR1 and E1 demonstrate minimal divergence throughout the study period.

B.1 Phylogenetic analysis



B.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample(868 days prior to week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 6 – blue, Week 4 – pink, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree has been affected by the inclusion of the retrospective sample.B.1 Subpopulation analysis

```
                        1        10         20      27
                        |--------+---------+------|
1209-0-18_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPHQK
            8_M13uni    ATYTTGGAQAYNTFRLTSLFDSGPQQK
            3_M13uni    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-0-8_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-0-1_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
      1160418_M13uni    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106617_M13uni106-6-    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-6-4_M13uni_--_2    ATYTTGGAQAYNTFRLTSLFDSGPQQK
110660-16-1310660-16    ATYTTGGAQAYNTFRLTSLFDSGPQQK
          110660-16-2   ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1410_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1219_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-6-5_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1213_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-123_M13uni_--_5    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-1210_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-126_M13uni_--_9    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-122_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-121_M13uni_--_1    ATYTTGGAQAYNTFRLTSLFDSGPQQK
1106-6-6_M13uni_--_6    ATYTTGGAQAYNTFRLTSLFDSGPQQR
       110666_M13uni    ATYTTGGAQAYNAFRLTSLFDSGPQQK
1106-1214_M13uni_--_    ATYTTGGAQAYNAFRLTSLFDSGPQQK
            11_M13uni   ATYTTGGAQAYNTFRLTSLFDPGPQQK
1106-0-14_M13uni_--_    ATYTTGGAQAYNTFRLTSLFDPGPQQK
L-1_M13uni-21_--_11.    ATYTTGGAQAYTTFRITSLFDSGPQQK
L-14_M13uni-21_--_10    ATYTTGGAQAYTTFRLTSLFDSGPQQK
L-12_M13uni-21_--_22    ATYTTGGAQAYTAFRLTSLFDSGPQQK
L-15_M13uni-21_--_11    ATYTTGGAQAYTAFRLTSLFDSGPQQK
L-18_M13uni-21_--_13    ATYTTGGAQAYTAFRLTSLFDSGPQQK
         110660-16-12   ATYTTGGAQAYHAFRLTSLFDSGPQQK
1106-1220_M13uni_--_    ATYATGGTQAYNTFRLASLFDSGPPQK
1106-1211_M13uni_--_    ATYATGGTQAYNTFRLVSLFDSGPPQK
         110660-16-20   ATYATGGSDAYDTFRLTSLFDSGAQQK
L-2_M13uni-21_--_12.    ATYTTGGAQAHTAARLTSLFDLGPQQK
          110660-16-5   ATYTTGGSQAHTTSRFTSFFDLGPQQK
            Consensus   ATYtTGGa#Ay.tfRltSlFDsGpqQK
```

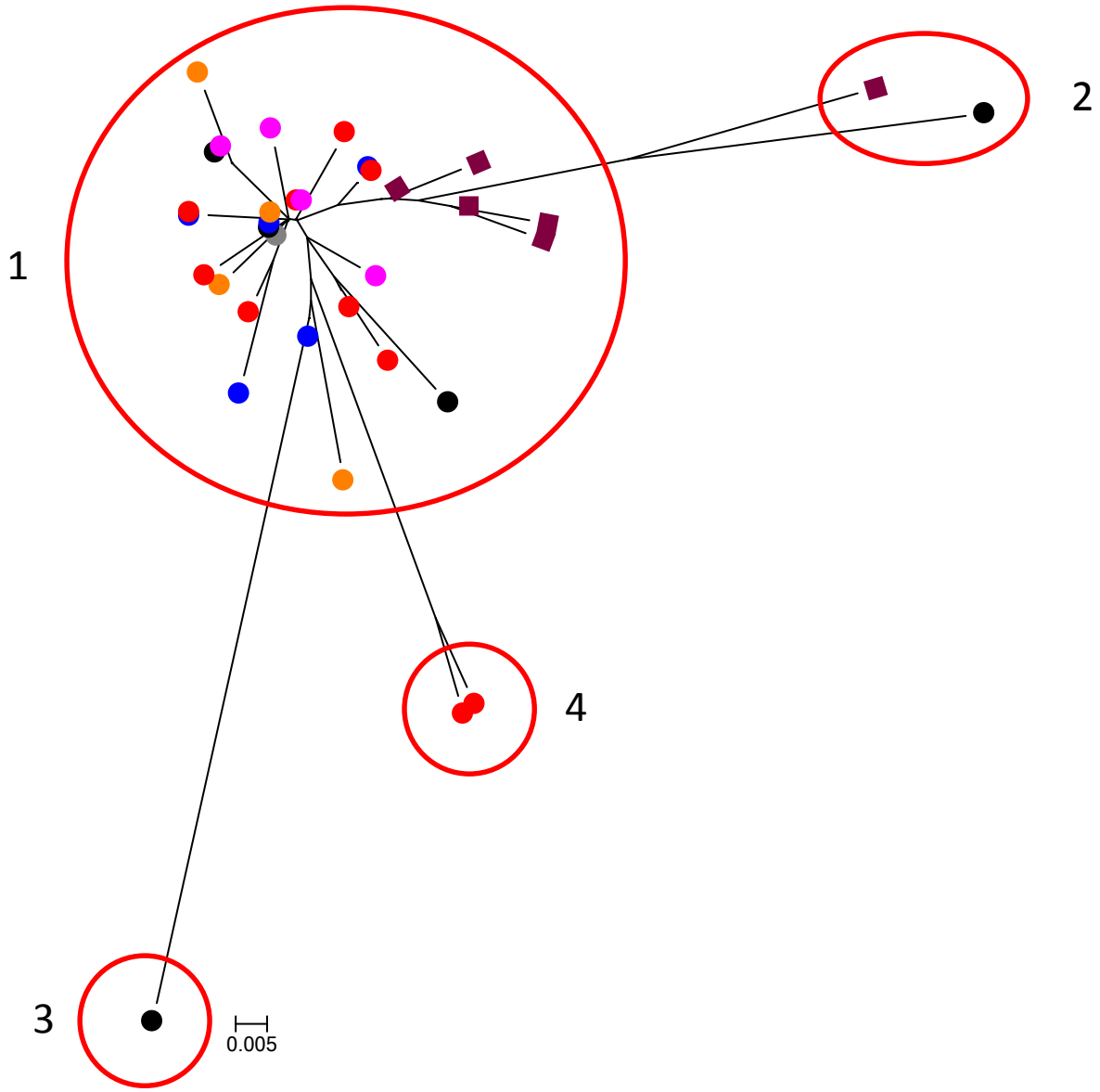
1

4

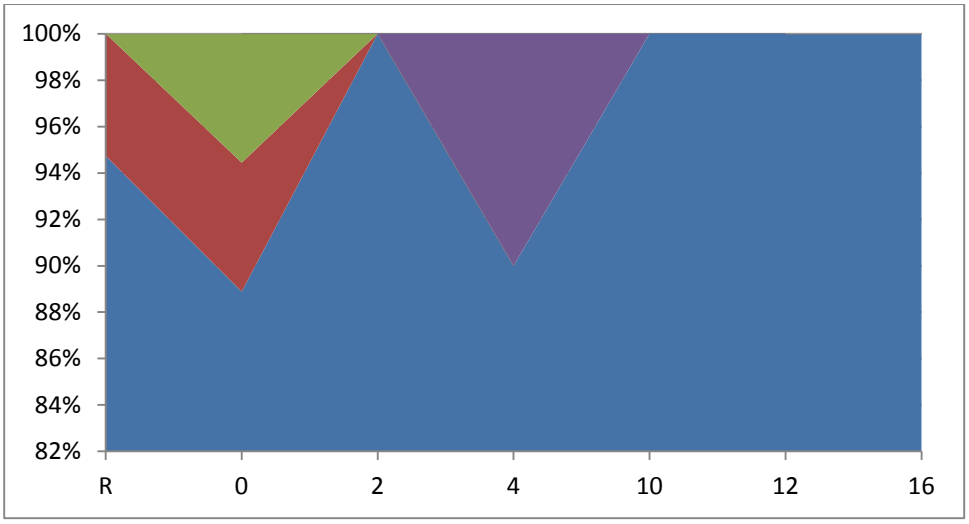
3

2

B.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



B.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 6 – blue, Week 4 – pink, Week 0 – orange.

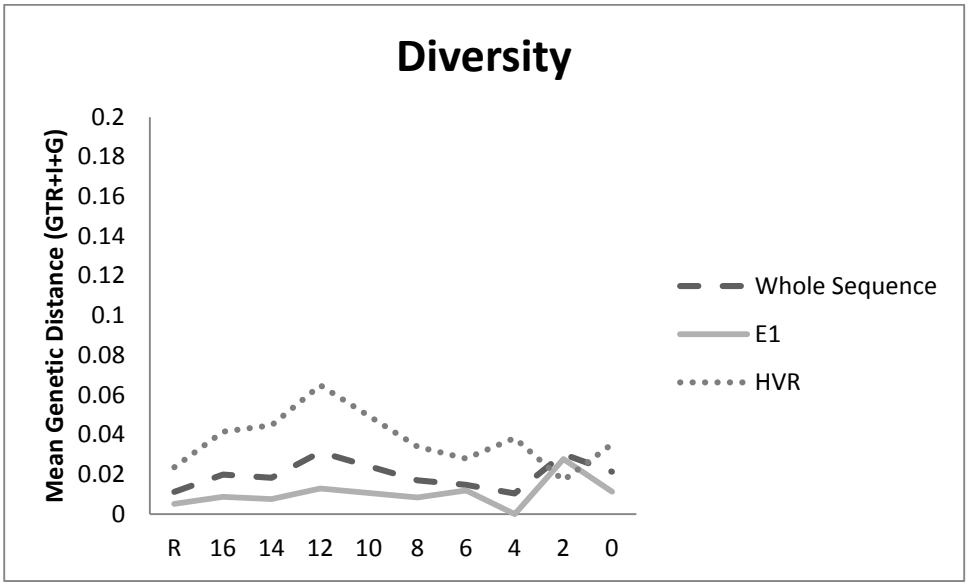


B.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subpopulation 1 comprises >90% of the clones sequenced throughout the study and including the retrospective samples.

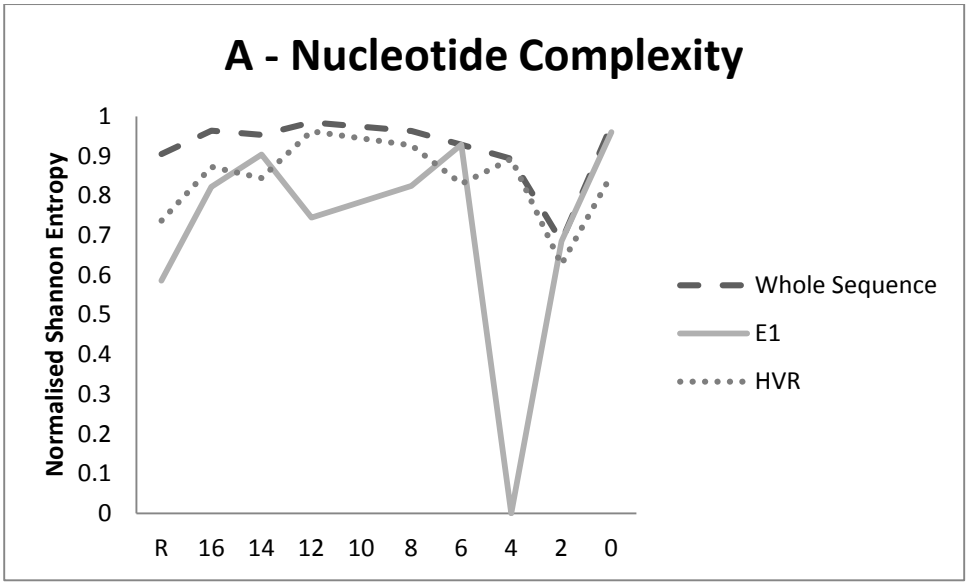
D.1 Subject D

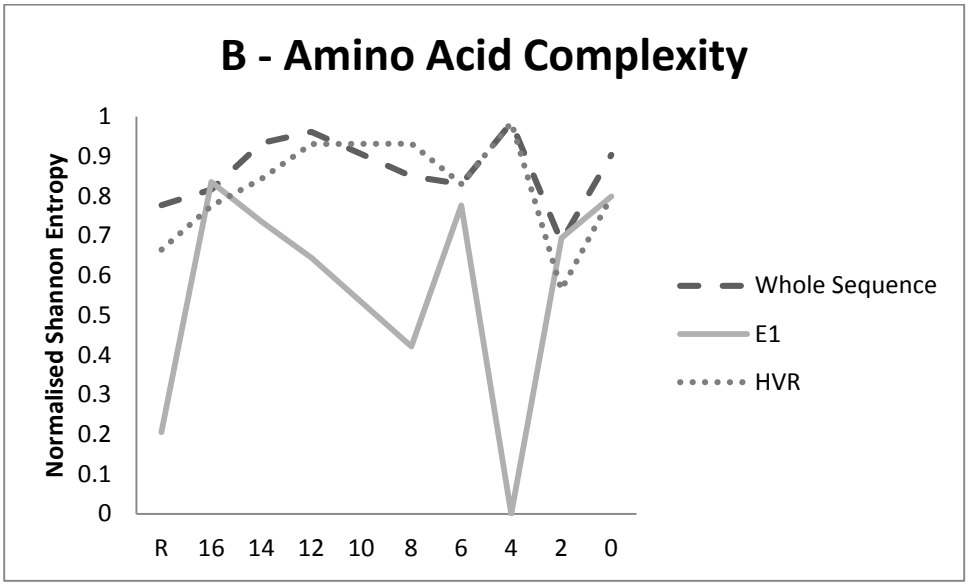
D.1 Diversity, Complexity, and Divergence



D.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

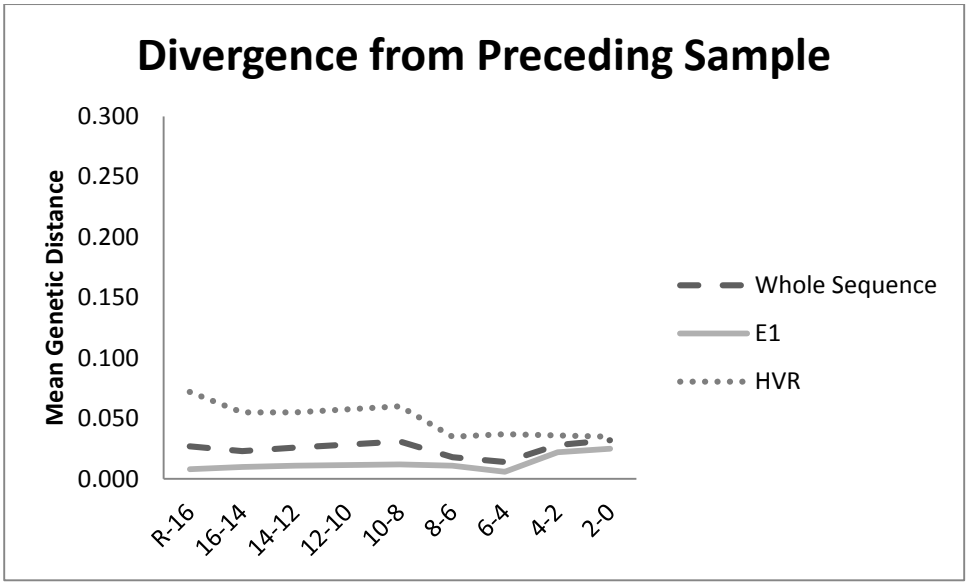
HVR1 diversity is greater than E1 diversity in most samples though there is a marked reduction in HVR1 QS diversity between weeks 12 and 6 which may suggest episodic selection with homogenisation of the QS milieu.





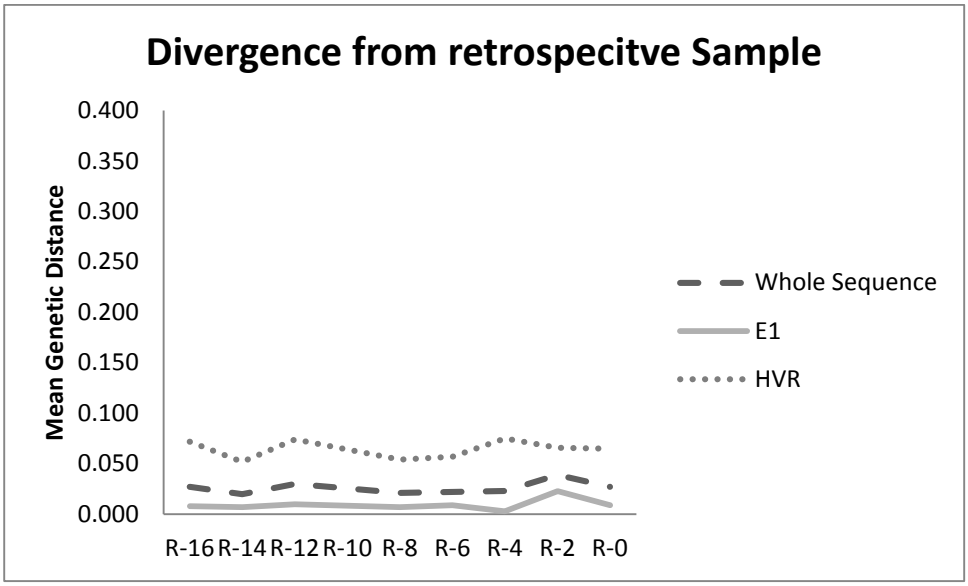
D.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased complexity relative to E1 at both nucleotide and amino acid level in most samples.



D.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

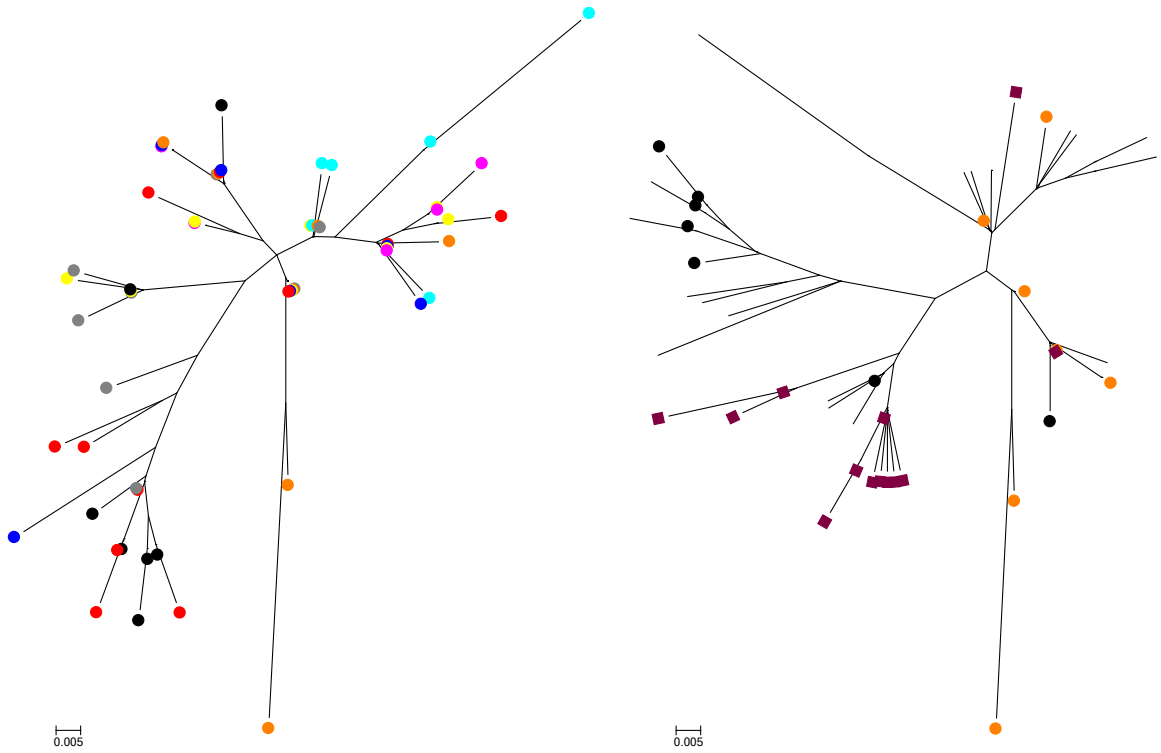
E1 divergence is minimal. HVR1 divergence is maximal between the retrospective sample and the week 16 sample though the inter sample divergence for the remainder of the study is comparable.



D.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group is negligible for the remainder of the study.

D.1 Phylogenetic analysis



D.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (252 prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.

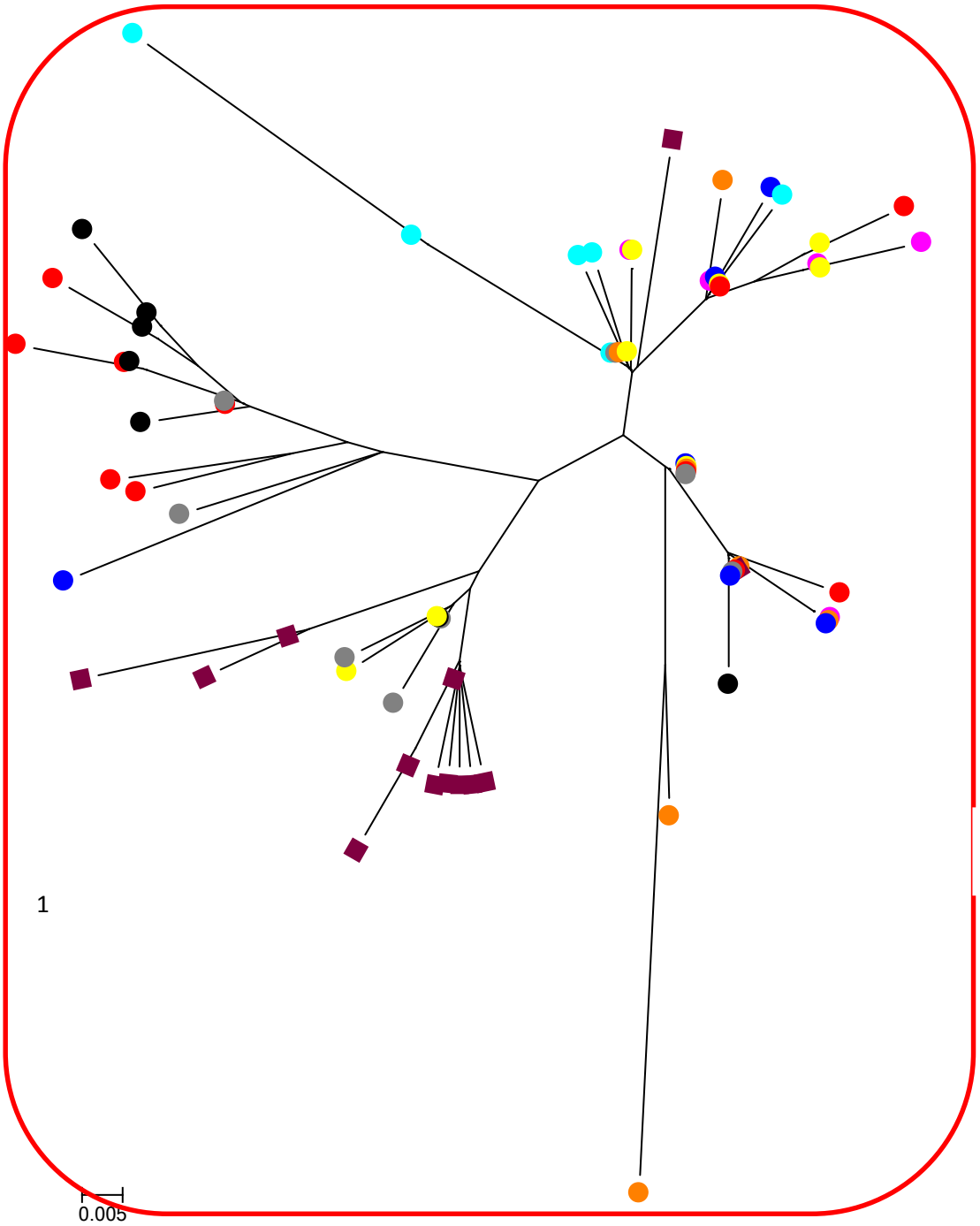
It is noticeable that the general shape of the tree has been affected by the inclusion of the retrospective sample. The result is some flipping of the clades in the top of the tree but there is little substantive change to the positioning of the different sequences.

```
                         1        10         20      27
                         |--------+---------+------|
1209-0-13_M13uni_--_     HTYMTGGVPSRSAYQLTSLFTSGARQN
1209-6-10_M13uni_--_     HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-8-7_M13uni_--_1     HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-1212_M13uni_--_     HTYVTGGVPSRSAYQLTSLFTSGARQN
1209-0-12_M13uni_--_     HTYVTGGVPSRSAYQLTSLFTSGARQN
Q-14-10_M13uni_--_12     HTYVTGGVPSRSAYQLTSLFTSGARQN
Q-14-14_M13uni_--_27     HTYVTGGVPARSAYQLTSLFTSGARQN
1209-4-11_M13uni_--_     HTYVTGGVPSRSAHQLASLFTSGARQN
1209-6-21_M13uni_--_     HTYVTGGVPSRSAHQLASLFTSGARQN
             1209-0-6    HTYVTGGVPSRSAHQLASLFTSGARQN
1209-6-17_M13uni_--_     HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-121_M13uni_--_3     HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-1218_M13uni_--_     HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-0-7_M13uni_--_1     HTYVTGGVPSRSAHQLTSLFTSGARQN
Q-14-1_M13uni_--_9..     HTYVTGGVPSRSAHQLTSLFTSGARQN
Q-10_M13uni-21_--_11     HTYVTGGVPSRSAHQLTSLFTSGARQN
1209-1612_M13uni_--_     NTYVTGGVPSRSAHQLTSLFTSGARQN
1209-2-1_M13uni_--_2     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-2-3_M13uni_--_3     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-1_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-10_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-4_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-4-7_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-6-11_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-2_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
Q-14-5_M13uni_--_27.     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-4_M13uni_--_9     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-0-2_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-125_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-6_M13uni_--_1     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-8a_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-1217_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-8-15_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-0-11_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARQN
1209-2-10_M13uni_--_     HTYVTGGVQSRSAFQLTSLFTSGARQN
1209-6-3_M13uni_--_1     HTYVTGGVQSRSTYQLTSLFTSGARQN
1209-2-12_M13uni_--_     HTYVTGGVQSRSAYQLTSLFTSGARHN
1209-2-2_M13uni_--_1     HTYVTGGVQSRSAYQLMSLFTSGARQN
1209-2-6_M13uni_--_3     HTYVTGGVQSRSASQLTSLFTSGARQN
1209-1210_M13uni_--_     HTYVTGGVQARSAHQLTSLFTSGARQN
1209-1215_M13uni_--_     HTHLTGGVQARSAHQLTSLFTSGARQN
1209-0-1_M13uni_--_1     HTYMTGGAPSRSAYQLASLFTSGARQN
1209-1213_M13uni_--_     HTHVTGGAQARSAYQLTSLFTSGAGQN
1209-1216_M13uni_--_     HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1219_M13uni_--_     HTHVTGGAQARSAYQLTSLFTSGARQN
1209-8-8_M13uni_--_1     HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1619_M13uni_--_     HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1617_M13uni_--_     HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1618_M13uni_--_     HTHVTGGAQARSAYQLTSLFTSGARQN
1209-1220_M13uni_--_     HTHLTGGAQARSAYQLTSLFTSGARQN
1209-1610_M13uni_--_     HTHLTGGAQARSAYQLTSLFTSGARQN
1209-161_M13uni_--_1     HTHLTGGAQARSAYQLTSLFTSGARQN
Q-13_M13uni-21_--_7.     HTHVTGGAQSRSAYQLTSLFTPGARQN
1209-6-13_M13uni_--_     RTHVTGEAPARSAYQLTSLFTSGARQN
1209-8-3_M13uni_--_1     RTHVTGGVPSRSAYQLTSLFTSGARQN
Q-14-16_M13uni_--_26     RTHVTGGVPSRSAYQLTSLFTSGARQN
Q-14-24_M13uni_--_10     RTHVTGGVPSRSAYQLTSLFTSGARQN
Q-14-7_M13uni_--_67.     RTHVTGGVPSRSAYQLTSLFTSGARQN
1209-1620_M13uni_--_     RTHVTGGVPSRSAYQLTSLFTSGARQN
1209-8-11_M13uni_--_     RTHVTGGVPSRSAYQLTSLFTSGARQN
Q-1_M13uni-21_--_9..     RTHVTGGAPSRSAYQLTSLFTSGARQN
Q-16_M13uni-21_--_11     RTHVTGGAPSRSAYQLTSLFTSGARQN
Q-14_M13uni-21_--_21     RTHVTGGAPSRSAYQLTSLFTSGARQN
Q-15_M13uni-21_--_10     RTHVTGGAPSRSAYQLTSLFTSGARQN
Q-12_M13uni-21_--_18     RTHVTGGAPSRSVYQLTSLFTSGARQN
Q-10_M13uni-21_--_12     RTHVTGGAPSRSAYQLTSLFTPGARQN
Q-22_M13uni-21_--_10     RTHVTGGAPSRSAYQLTSLFTPGARQN
Q-24_M13uni-21_--_21     RTHVAGGAPSRSAYQLTSLFTSGARQN
Q-16_M13uni-21_--_10     RTHVTAGAQSRSAYQLTSLFTPGARQN
Q-18_M13uni-21_--_11     RTHVTAGAQSRSAYQLTSLFTSGARQN
Q-24_M13uni-21_--_10     RTHVTAGAQSRSAYQLTSLFTSGARQN
                Consensus hThvTGGvqsRSAyQLTSLFTSGARQN
```

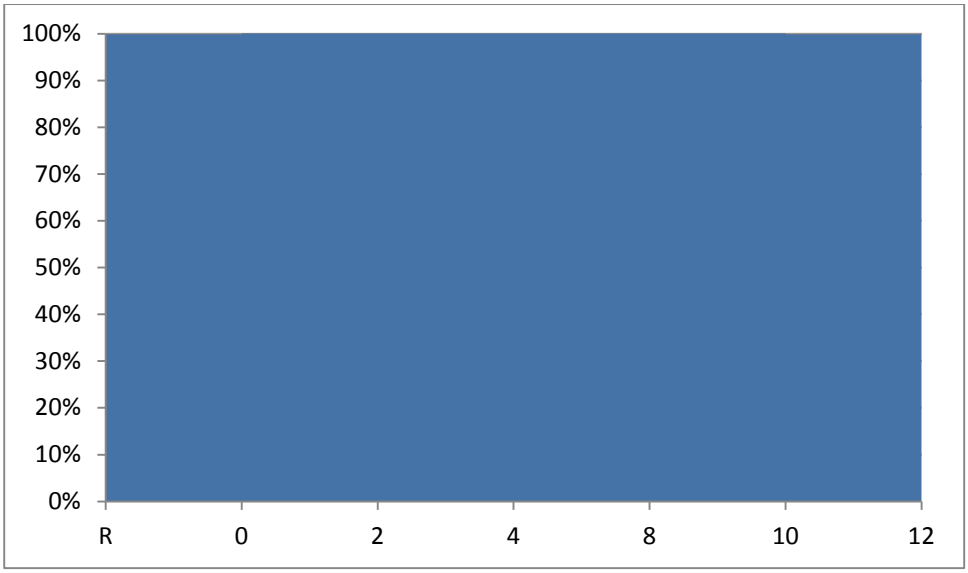
1

2

D.1 Fig 7 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (2 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which included the greater number of sequences was labelled first.



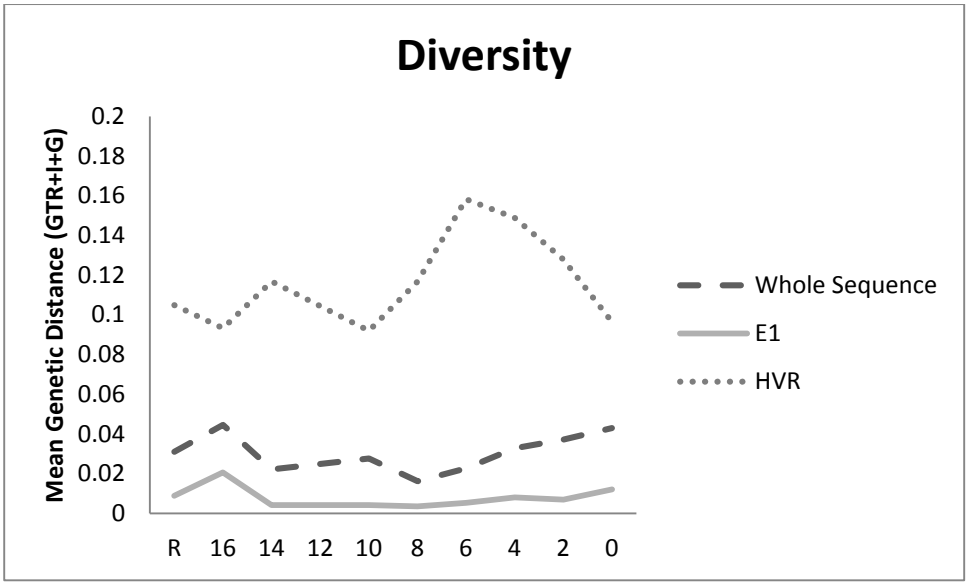
D.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulation as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange.



B.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

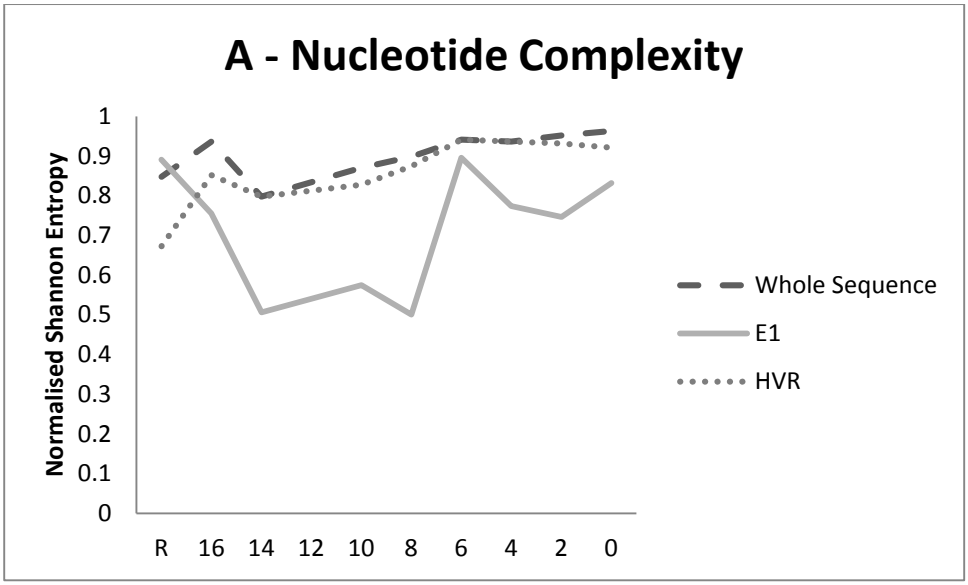
G.1 Subject G

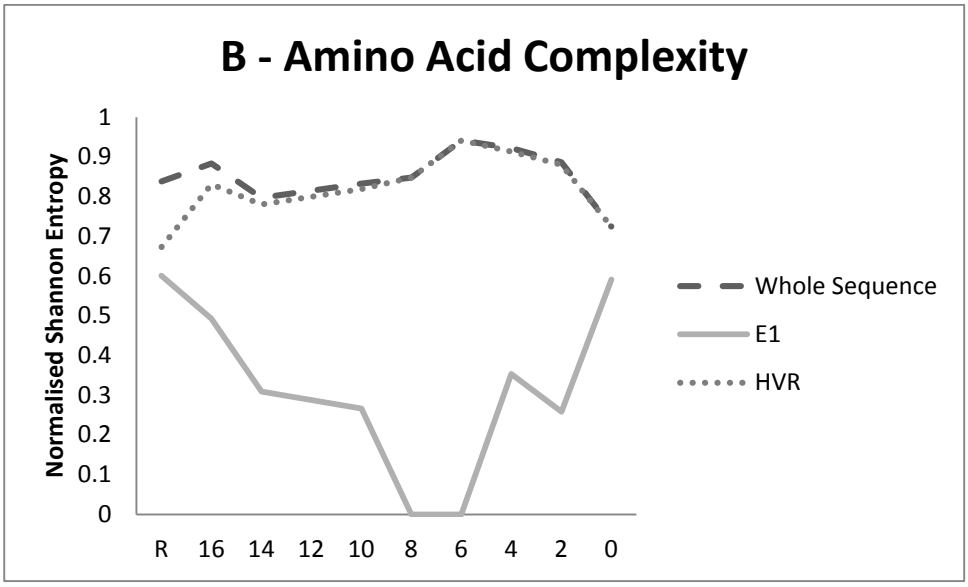
G.1 Diversity, Complexity, and Divergence



G.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

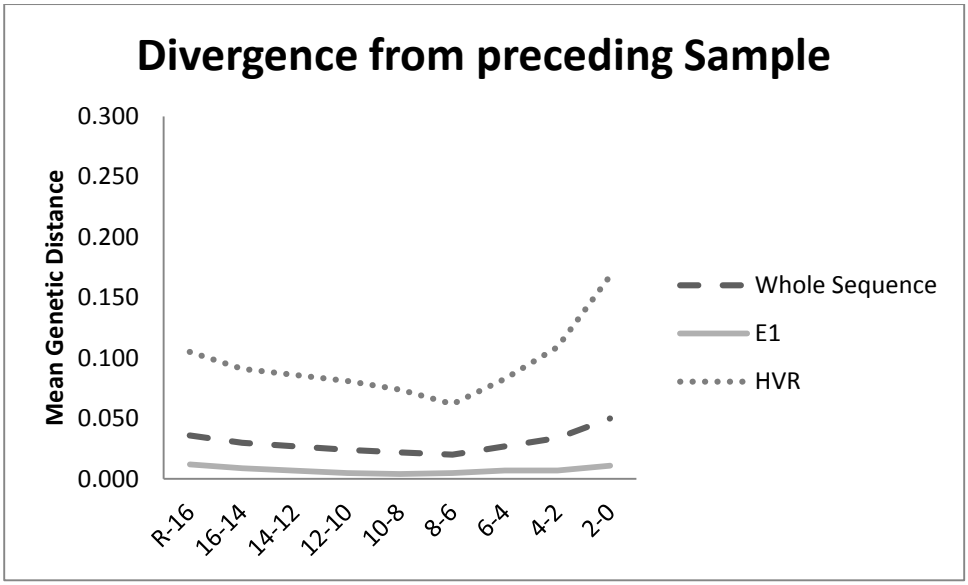
HVR1 diversity is far greater than E1 diversity.





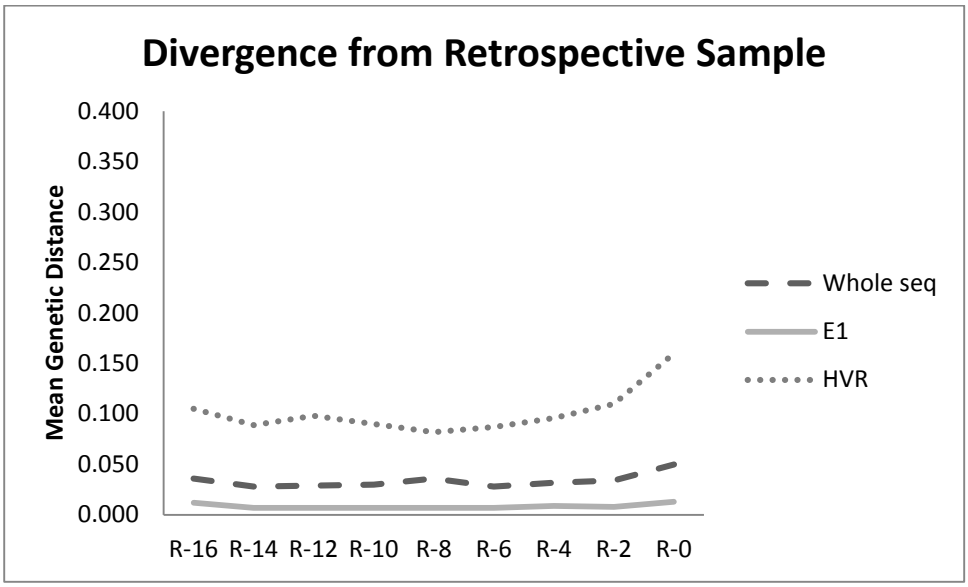
G.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates greater complexity relative to E1 throughout the study period.



G.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

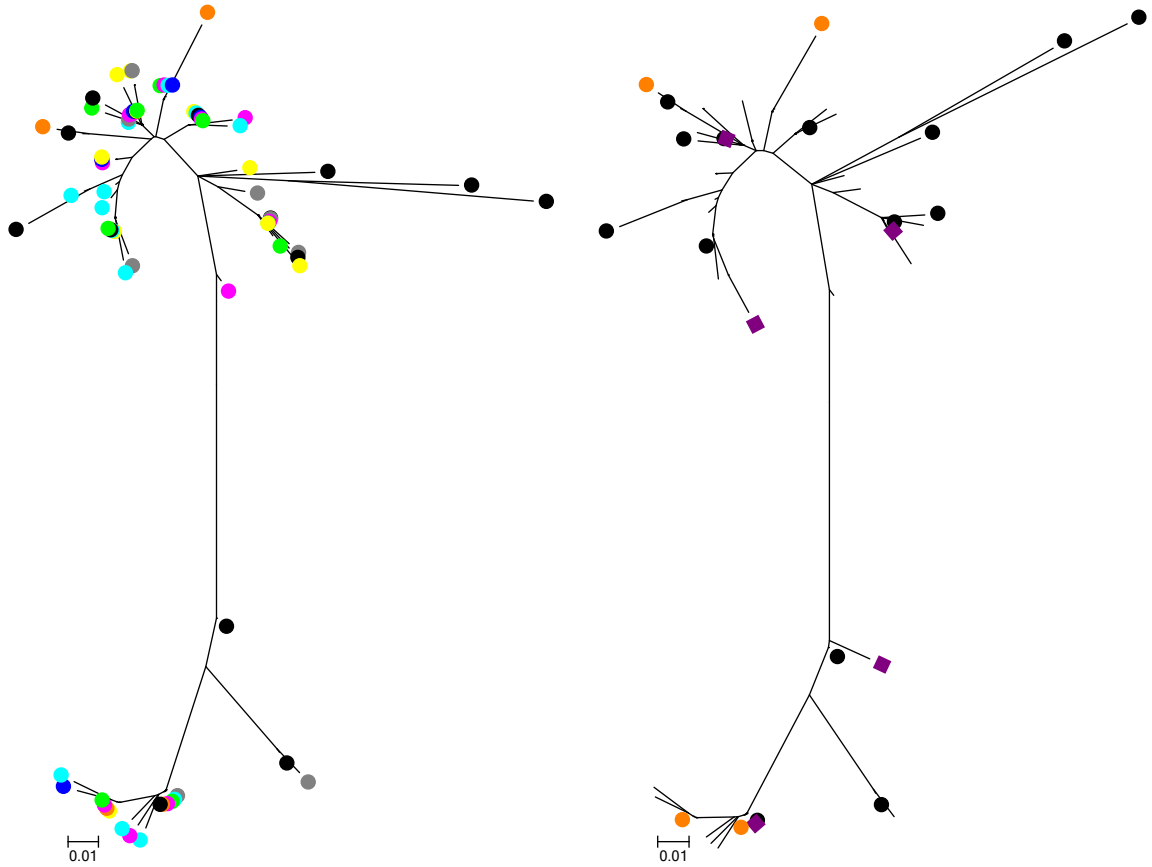
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that there is a similar magnitude of divergence.



G.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. The divergence between the retrospective group of clones is maximal when compared with the sample taken immediately pre treatment. This suggests ongoing divergent change potentially indicating a virus under diversifying selective pressures.

G.1 Phylogenetic analysis



G.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (122 dyas prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

Here we see that the retrospective samples do not materially alter the phylogeny. This suggests that evolution of HVR1 sequences has been confined to local fitness maxima.

G.1 Subpopulation analysis

```
                          1         10        20      27
                          |--------+---------+------|
                2012832   TTHTIGGAAAHGTRALTSLFSLGAQQK
                2012-4-7  TTYTTGGAAARGTHALTILFSLGAQQK
        A-6-18-6-3-6-6    TTYTTGGAAARGTHALTSLFSLGAQQK
A-4-19-4-10-4-13-4-1      TTYTTGGAAARGTHALTSLFSLGAQQK
       2012-818012-8-7    TTYTTGGAAARGTHALTSLFSLGAQQK
               2012168    TTYTTGGAAARGTHALTSLFSLGAQQK
                 A-2-12   TTYTTGGAAARGTHALTSLFSLGAQQK
               2012-1014  TTYTTGGAAARGTHALTSLFSLGAQQK
                  A-2-7   TTYTTGGAAARGTHTLTSLFSLGAQQK
               2012-4-12  TTYTTGGAAAHGTHALTSLFSLGAQQK
A-6-10-6-13-6-15-6-1      TTYTTGGAAAHGTHALTSLFSLGAQQK
2012-8-1012-8-14012-      TTYTTGGAAAHGTHALTSLFSLGAQQK
               2012-8-12  TTYTTGGAAAHGTHALTSLFSLGAQQK
2012-1013012-1015012      TTYTTGGAAAHGTHALTSLFSLGAQQK
               2012-1023  TTYTTGGAAAHGTHALTSLFSLGAQQK
2012-141012-1419012-      TTYTTGGAAAHGTHALTSLFSLGAQQK
A-1-11-12-13-18-19-2      TTYTTGGAAAHGTHALTSLFSLGAQQK
               2012-1415  TTYTTGGAAAHGTHALTSLFSLGAQQK
20121617012-164012-1      TTYTTGGAAAHGTHALTSLFSLGAQQK
     A-2-10-2-20-2-21     TTYTTGGAAAHGTHALTSLFSLGAQQK
                2012-166  TTYTTGGAAAHGTHALTSLFSLGAQQK
            A-4-24-4-8    TTYTTGGAAAHGTHALTSLFSLGAQQK
               2012-8-6   TTYTTGGAAAHGTHALTSLFTLGAQQK
        A-6-1-6-14-6-9    TTYTTGGAAAHGTHVLTSLFSLGAQQK
               2012-8-10  TTYTTGGAAAHGTHVLTSLFSLGAQQK
                   A-4-9  TTYTTGGAAAHGTHVLTSLFSLGAQQK
               2012-8-13  TTYTTGGVAAHGTHALTSLFSLGAQQK
            A-6-11-6-5    TTYTTGGAAAHGTRALTSLFSLGAQQK
                   A-4-4  TTYTTGGAAAHGTRALTSLFSLGAQQK
 A-2-1-2-18-2-2-2-22      TTYTTGGAAAHGTRALTSLFSLGAQQK
               2012-1016  TTYTTGGAAAHGTRALTSLFSLGAQQK
                2012162   TTYTSGGSAAHGTHAITSLFTLGAQQK
      2012-1618012167     TTYTTGGAAAHRTHALTSLFSVGAQQK
                2012839   TTYTTGGAAAHRTHVLTSLFSVGAQQK
               2012-8-16  TTYTTGGAAAHGTHVLTRLFSQGAQQK
               2012-1024  TTYTTGGAAAHGTHVLTRLFSQGAQQK
     201216100121619      TTYTTGGAAAHGTHVLTRLFSQGAQQK
A-2-11-2-15-2-16-2-8      TTYTTGGAAAHGTHVLTRLFSQGAQQK
                 A-2-9    TTYTTGGAAAHGTHVLTRLFSQGAQQK
                   A-15   TTYTTGGAAAHGTHMLTRLFSQGAQQK
               2012-1414  TTYTTGGAAAHGTHMLTRLFSQGAQQK
                 A-2-5    TTYTTGGAAAHGTHVLTRLFSPGAQQK
                2012-163  TTYTTGGAAAHGTHVFTRLFNLGAQQK
                 A-2-4    TTYTTGGAAAHGTHVFTRLFSLGAQQK
                 A-2-3    TTYTTGGAAAHGTHVLTRLFSLGAQQK
                2012-4-2  TTYTTGGAAARGTHALAGLFSLGAQQK
2012-8-6_dupName012-      TTYTTGGAAAHGTRALTRLFTLGAQQK
                A-4-15    TTYTTGGAAAHGTRALTRLFTLGAQQK
2012-1010012-1011012      TTYTTGGAAAHGTRALTRLFTLGAQQK
2012-1619012-1620012      TTYTTGGAAAHGTRALTRLFTLGAQQK
               2012-161   TTYTTGGAAAHGTRALTRLFTLGAQQK
                   A-10   TTYTTGGAAAHGTRALTRLFTLGAQQK
               2012-146   TTYTTGGAAAHGTRALTRLFTLGAQQK
2012-1413012-1416012      TTYTTGGAAAHGTRALTRLFTLGAQQK
               2012-1020  TTYTTGGAAAHGTRALTRLFTLGAQQN
                2012-8-8  TTYTTGGAAANGTRALTRLFTLGAQQK
                2012-144  TTYTTGGAAARGTHALTRLFTLGAQQK
               2012-1610  TTHITGGAAAHGTRALTSLFTPGAQQK
                2012-169  TTYITGGAAAHGTSALTSLFTLGAQQK
                2012-4-1  TTYTSGGSYARTTNAIAGLFTPGARQN
            A-6-17-6-4    TTYTSGGSAARTTNAIAGLFTPGARQN
                 A-2-6    TTYTSGGSAARTTNAIAGLFTPGARQN
                2012-8-2  TTYTSGGSAARTTNAIAGLFTPGARQN
                   A-4-2  TTYTSGGSAARTTNAIAGLFTPGARQN
20128330128340128311      TTYTSGGSAARTTNAIAGLFTPGARQN
      20121690121613      TTYTSGGSAARTTNAIAGLFTPGARQN
2012-101012-108012-1      TTYTSGGSAARTTNAIAGLFTPGARQN
                 A-2-19   TTYTSGGSAARTTNAIAGLFTPGARQN
               2012-1019  TTYTSGGSAARTTNAIAGLFTPGARQN
      20128360128318      TTYTSGGSAARTTNAIAGLFTPGARQN
          A-2-17-2-23     TTYTSGGSAARTTNAIAGLFTPGARQN
                 A-4-5    TTYTSGGSAARTTNAIAGLFTPGARQN
          A-2-13-2-14     TTYTSGGSAARTTNAIAGLFTPGARQN
               2012-145   TTYTSGGSAARTTNAIAGLFTPGARQN
          A-16-2-24-5     TTYTSGGSAARTTNAIAGLFTPGARQN
               2012-165   TTYTTGGAAAHTTNAIAGLFTPGARQN
                  A-22    TTYTTGGSAARTTNAIAGLFTLGAQQN
               2012-1421  TTYTSGGSAAHTTSLLAGLFNPGARQN
2012-161501216401216      TTYTSGGSAAHTTSLLAGLFNPGAQQN
              Consensus   TTYTtGGaAAhgT.alt.LF.lGAqQk
```

1

9

2

3

4

5

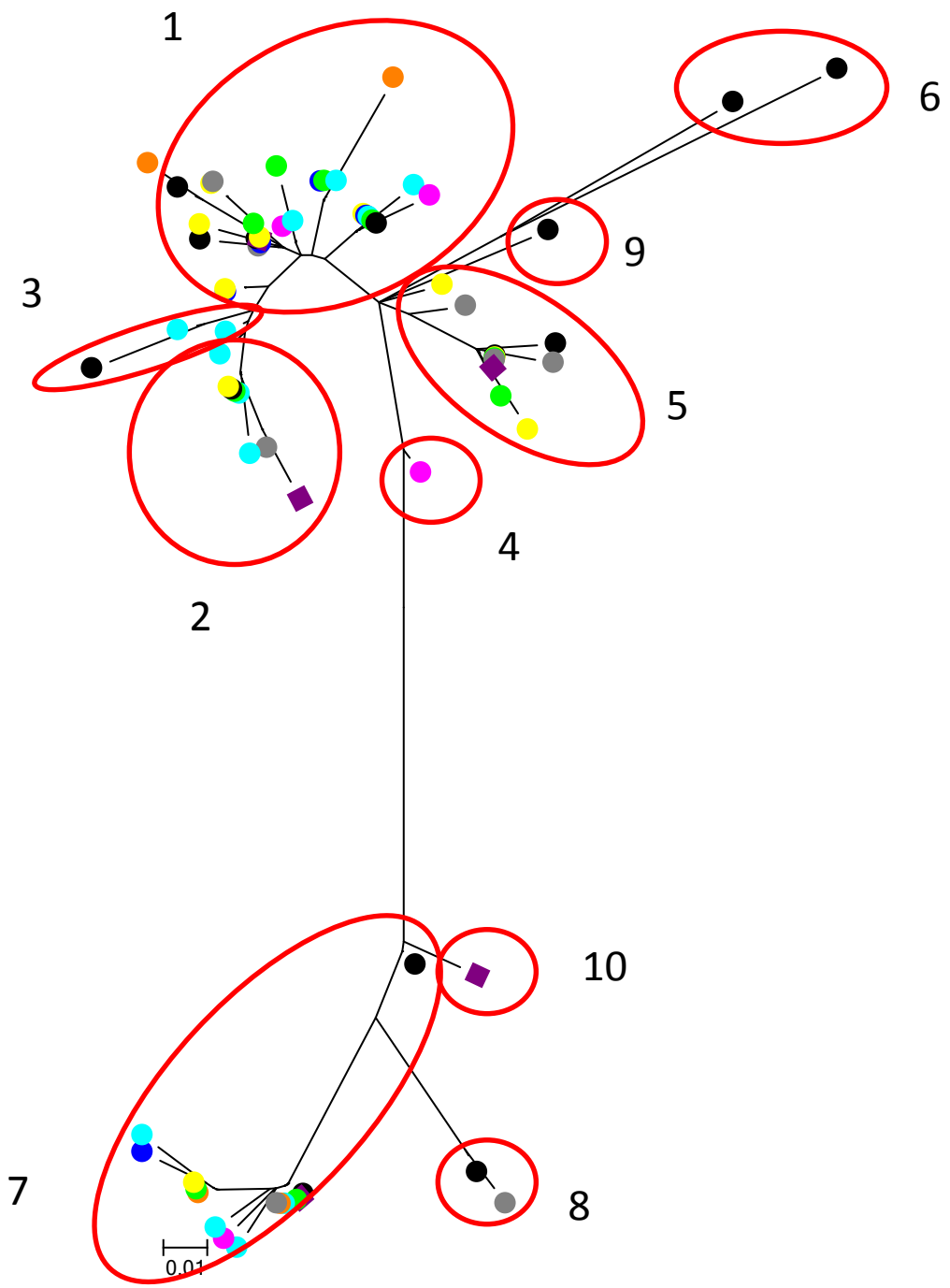
6

7

10

8

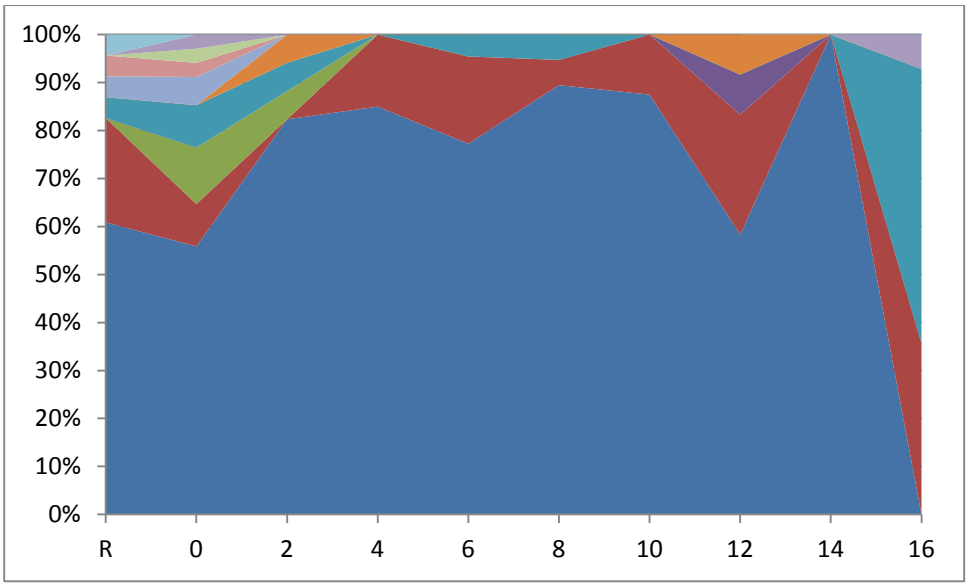
G.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



G.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

The use of subpopulation designations immediately identifies subpopulation 10 as an outlier. This subpopulation, not present in any of the subsequent clones generated may either represent an

ancestral master sequence or could also signify a previous effort to explore for alternative fitness benefits.



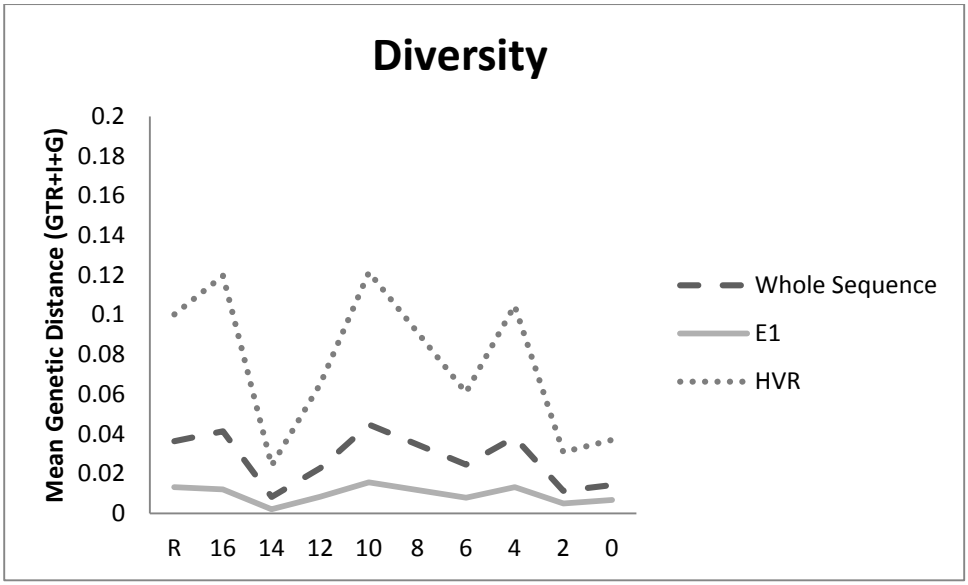
G.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

This illustration of subpopulation prevalence through the study period clearly identifies a dominant subpopulation which has been present for almost the entire 8 months prior to commencing treatment. This subpopulation has however been eliminated by the pre treatment sample suggesting immune mediated clearance.

PARRIS analysis of evidence for sequence wide positive selection is unaffected by the inclusion of retrospective samples.

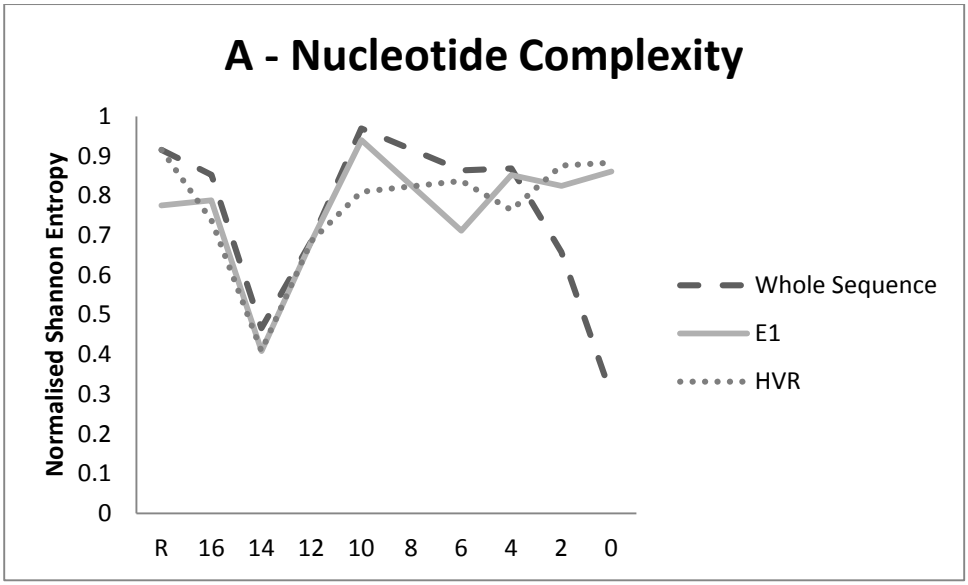
I.1 Subject I

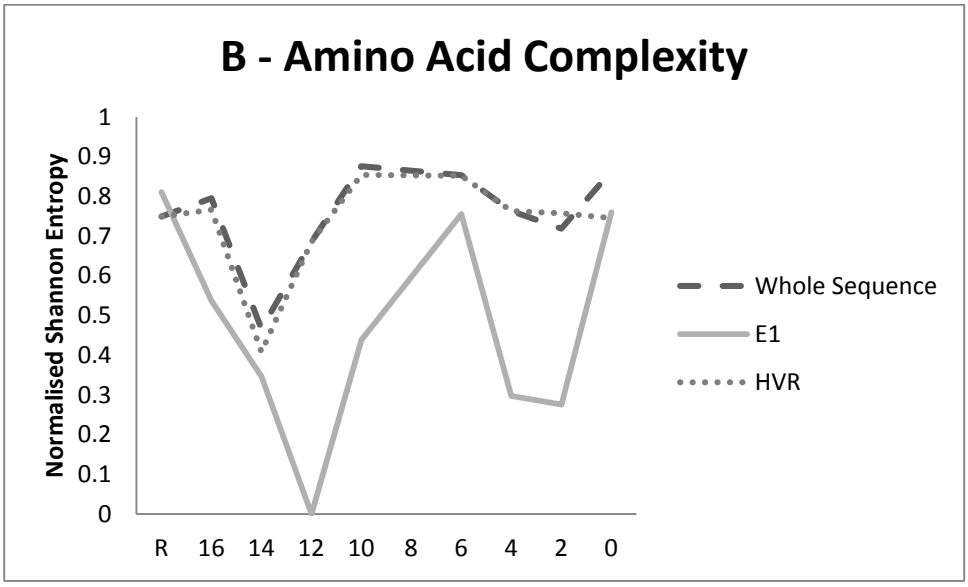
I.1 Diversity, Complexity, and Divergence



I.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

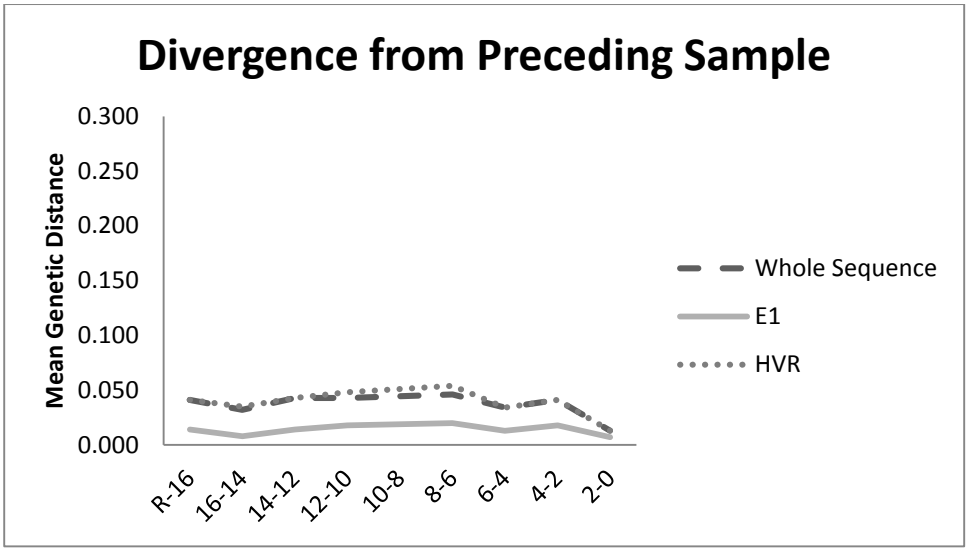
HVR1 diversity is far greater than E1 diversity.





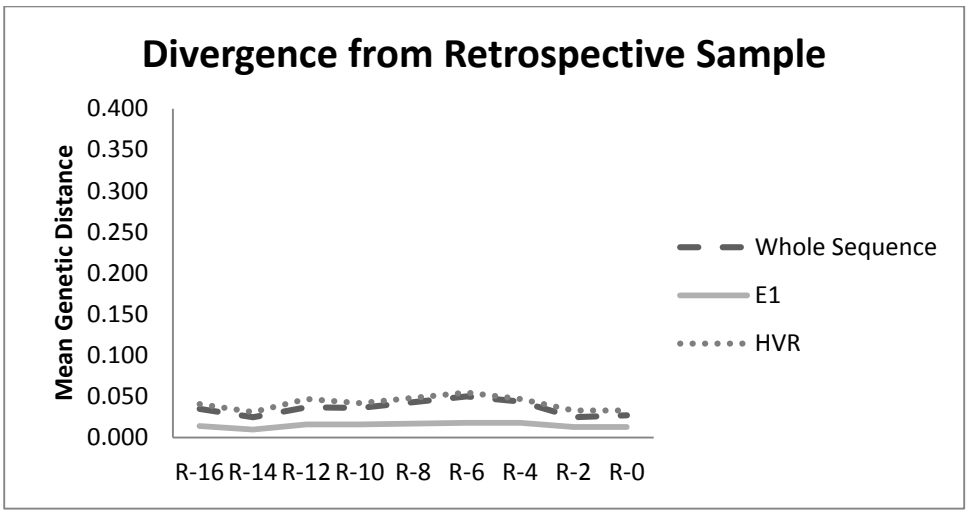
I.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased complexity relative to E1 at most points during the study period.



I.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

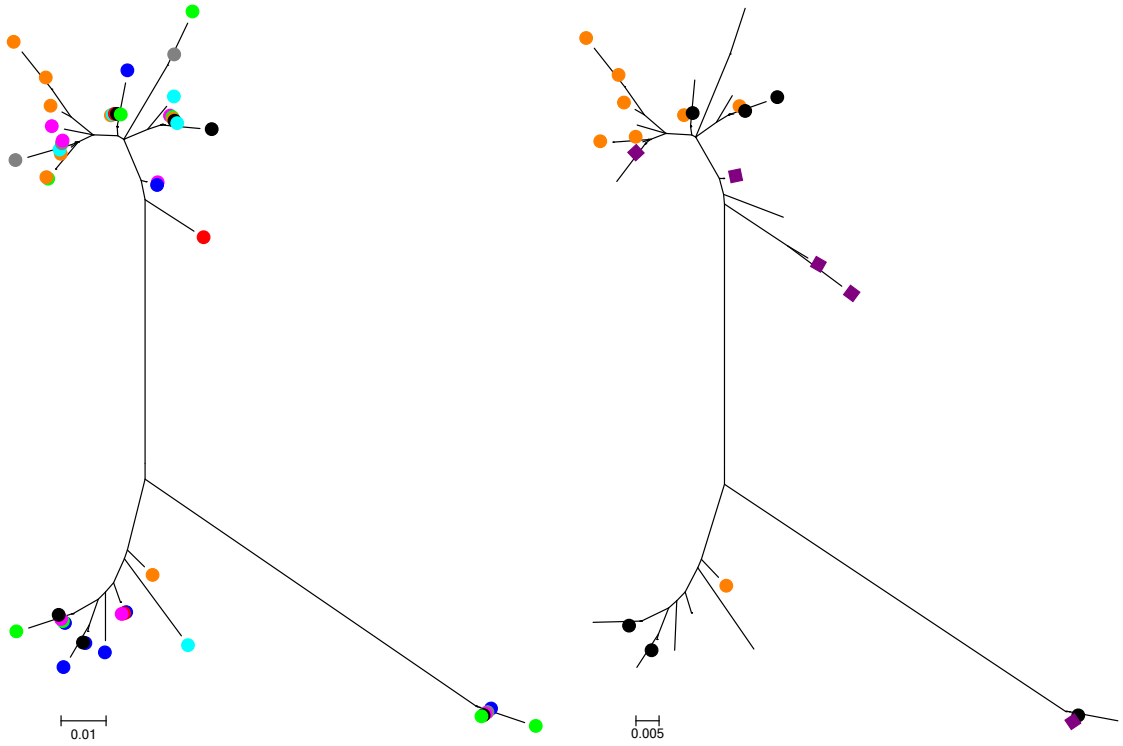
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that there is a similar magnitude of divergence.



I.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group appears maximal at timepoint 6 (6 weeks pre treatment). This potentially suggests that both divergent and convergent change in the HVR1.

I.1 Phylogenetic analysis



I.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (168 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree is unaffected by the inclusion of the retrospective sample.

I.1 Subpopulation analysis

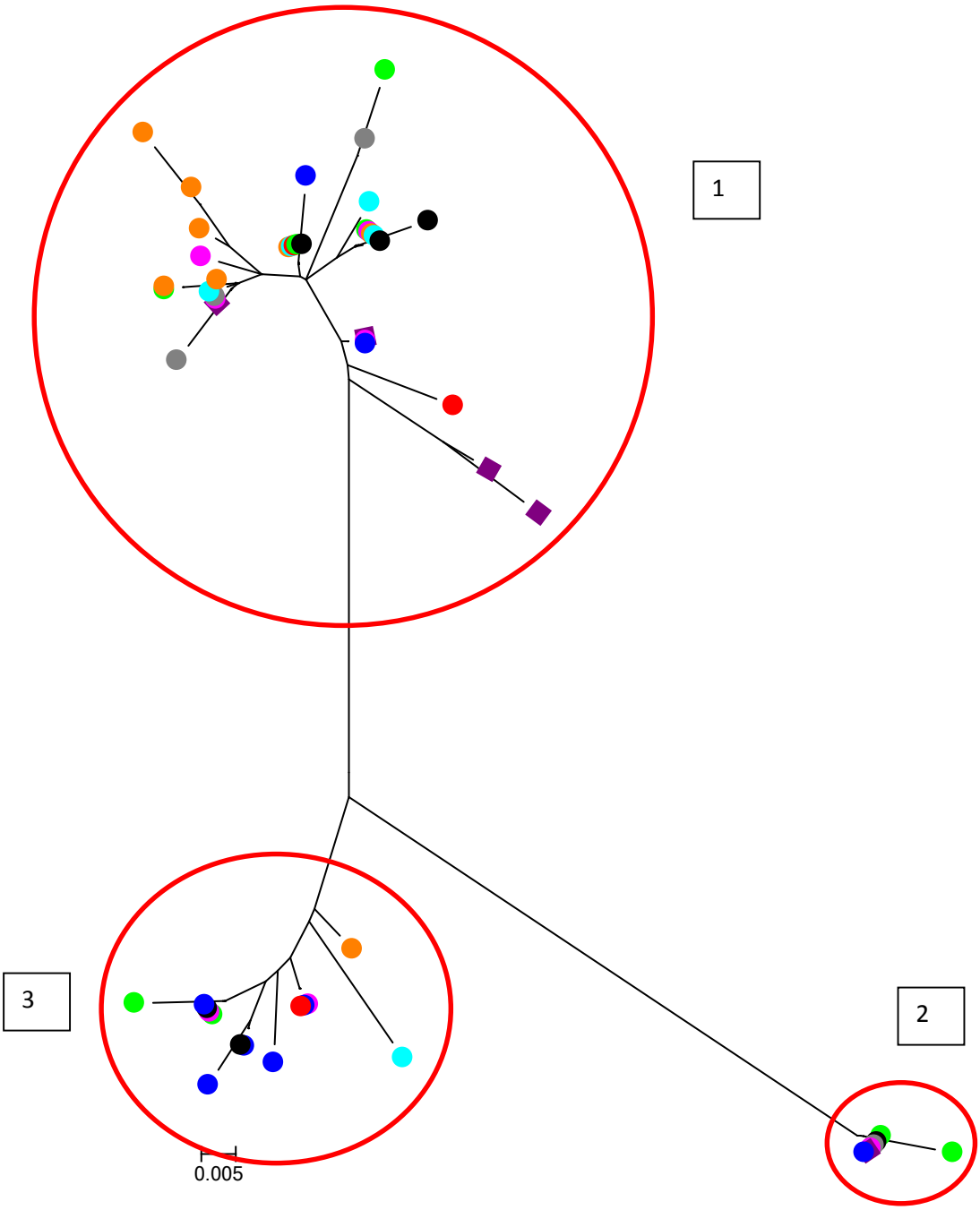
```
                          1        10        20        30        40        49
                          |--------+---------+---------+---------+--------|
           2607-0-20_                          TTHTIGGTMAHGVRSFSSLFSMGASQK
C-2-20_M13uni_--_7..                           TTHTIGGTMAHGVRSFSSLFSMGASQK
267-4-4_M13uni_--_13                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-6-13_M13uni_--_10.                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-20_M13uni-21_--_25                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-10-4_M13uni_--_10.                           TTHTIGGTMAHGVRSFSSLFSMGASQK
            2607-0-16                          TTHTIGGTMAHGVRSFSSLFSMGASQK
2607-124_M13uni_--_1                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-2-12_M13uni_--_18.                           TTHTIGGTMAHGVRSFSSLFSLGASQK
C-16-8_M13uni_--_24.                           TTHTIGGTMAHGVRSFSSLFSLGASQK
267-4-9_M13uni_--_15                           TTHTIGGTMAHGVRSFSSLFSLGASQK
2607-0-1_M13uni_--_1                           TTHTIGGTMAHGVRSFSSLFSLGASQK
C-10-19_M13uni_--_12                           TTHTIGGTMAHGVRTFSSLFSMGASQK
2607-0-3_M13uni_--_1                           TTHTIGGTMAHGVRTFSSLFSMGASQK
267-4-19_M13uni_--_1                           TTHTIGGTMAHGVRSFSSLFNMGASQK
            2607-1212                          TTHTIGGTMAHGVRSFSSLFSVGASQK
267-14-19_M13uni_--_                           TTYTIGGTMAHGVRSFSSLFSMGASQK
C-10-7_M13uni_--_82.                           TTHTIGGTMAHGVRSLSSLFSVGASQK
267-14-7_M13uni_--_1                           TTHTIGGTMAHGVRSLSSLFSVGASQK
C-16-9_M13uni_--_16.                           TTHTIGGTMARGVRSFSSLFSLGASQK
C-12_M13uni-21_--_38                           TTHIIGGTMAHGVSSFSSLFSMGASQK
C-2-1_M13uni_--_22..                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-10-1_M13uni_--_21.                           TTHTIGGTMAHGVRSFSSLFSLGASQK
267-4-1_M13uni_--_11                           TTRIIGGQAAHGVRTFSSLFSVGASQK
2607-1210_M13uni_--_                           TTRIIGGQAAHGVRTFSSLFSVGASQK
C-6-14_M13uni_--_13.                           TTRIIGGQAAHGVRTFSSLFNVGASQK
C-16-14_M13uni_--_11                           TTRIIGGEVARGASSFYGLFNMGASQK
267-14-1_M13uni_--_1                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-16-1_M13uni_--_9..                           TTHTIGGTMAHGVRSFSSLFSMGASQK
C-2-2_M13uni_--_10..                           TTRIIGGQAAHGVRSLSSLFSVGASQK
267-4-16_M13uni_--_1                           TTRIIGGQAAHGVRTFSSLFNVGASQK
C-16-10_M13uni_--_75                           TTRIIGGQAAHGVRTFSSLFNVGASQK
C-10-11_M13uni_--_21                           TTRIIGGQAAHGVRTFSSLFNVGASQK
C-6-1_M13uni_--_43..                           TTRIIGGQAAHGVRSFSSLFNVGASQK
C-6-10_M13uni_--_47.                           TTRIIGGQAAYGVRTFSSLFNVGASQK
C-6-11_M13uni_--_27.                           TTRIIGGQAAYGVRSFSSLFNVGASQK
267691616_M13uni_--_                           TTRIIGGQAAYGVRSFSSLFNVGASQK
C-6-12_M13uni_--_46.                           TTRIIGGQAAHGVRTFSSLFSVGASQK
2607-0-19_M13uni_--_                           TTRIIGGQTAHGVRTFSSLFSVGASQK
C-6-23_M13uni_--_19.                           TTRIIGGEVARGASSFYGLFNMGASQK
267-14-10_M13uni_--_                           TTRIIGGEVARGASSFYGLFNMGASQK
C-16_M13uni-21_--_20                           TTRIIGGEVARGASSFYGLFNMGASQK
C-10-18_M13uni_--_12                           TTRIIGGEVARGASSFYGLFNMGASQK
267-4-20_M13uni_--_1                           TTRIIGGEVARGASSFYGLFNMGASQK
           Consensus                  .......  .TThtIGGtmAhGvrsFssLFsmGASQK
```

1

2

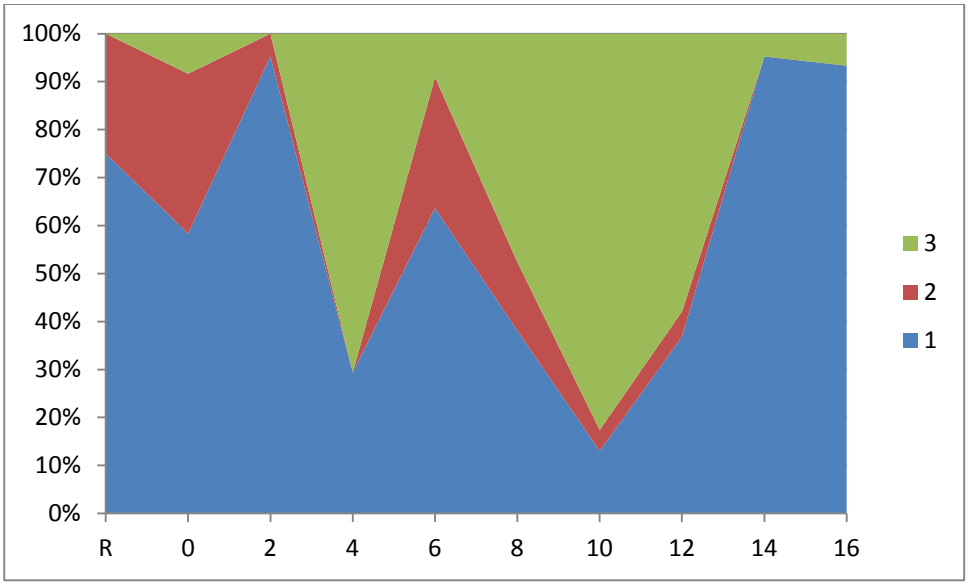
3

I.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



I.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

Subpopulation 3 is not identified in the retrospective sample suggesting that this has emerged in the intervening period. No new subpopulations were found in the retrospective sample.

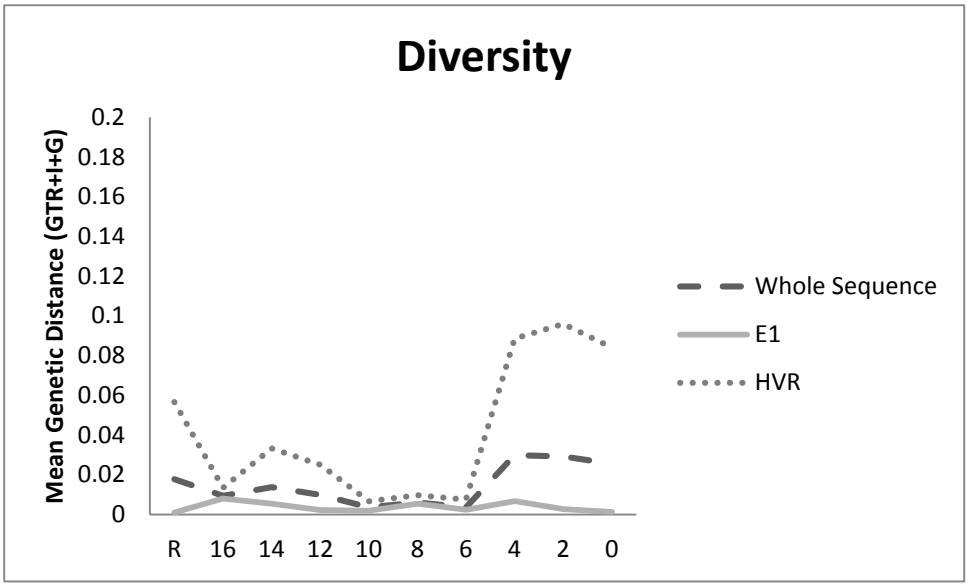


I.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subpopulation 3 as identified in Fig 7. Appears in the week 16 sample and transiently becomes the most dominant sequence prior to the pre treatment sample. Examination of the phylogeny of subpopulation 1 shows intra subpopulation change in the HVR1 QS suggesting that these sequence changes may have allowed this subpopulation to regain dominance within the entire HVR1 QS.

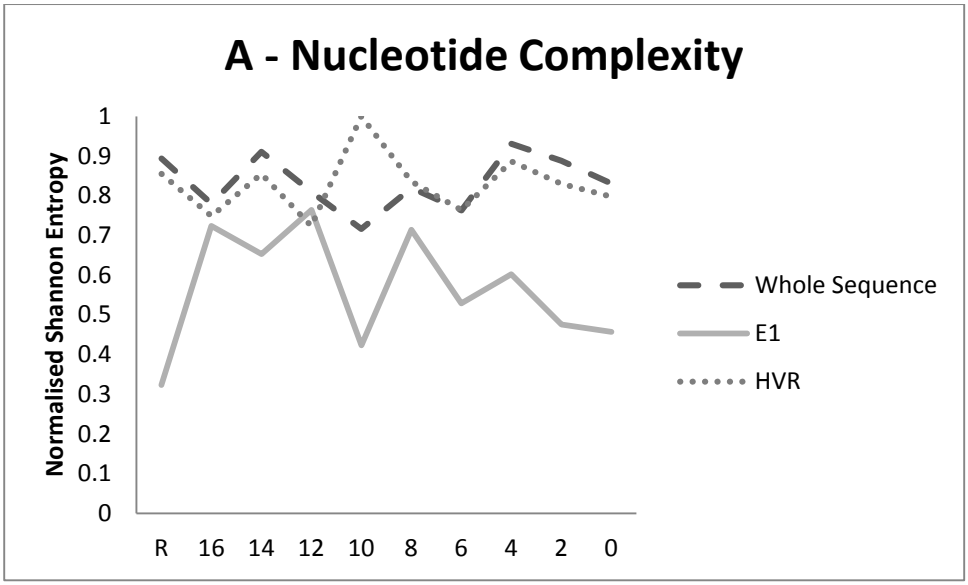
J.1 Subject J

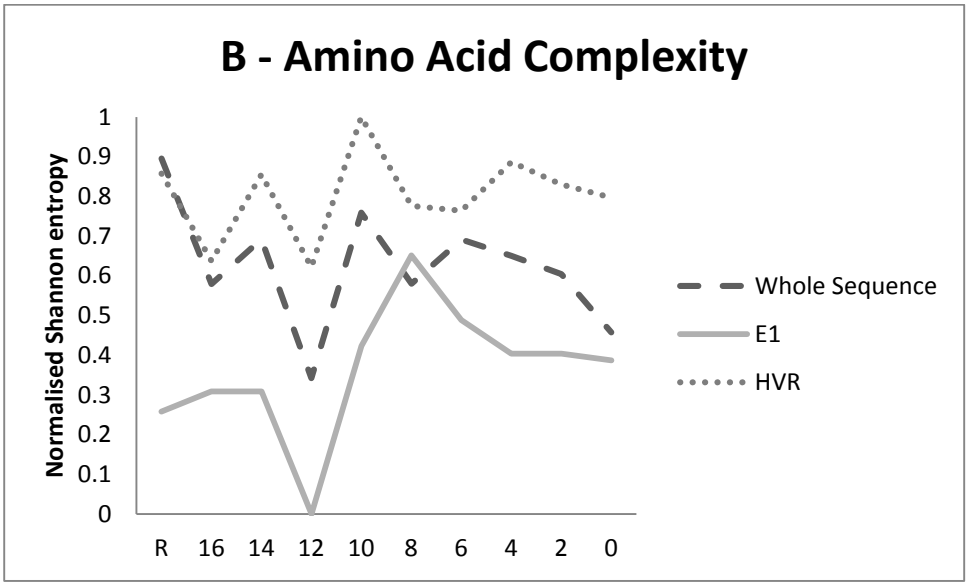
J.1 Diversity, Complexity, and Divergence



J.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

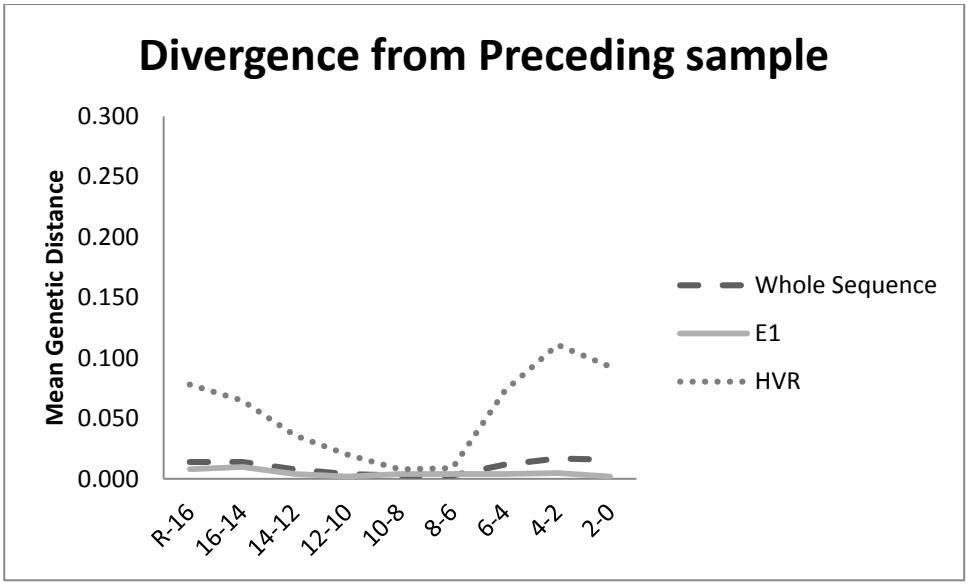
HVR1 diversity is far greater than E1 diversity.





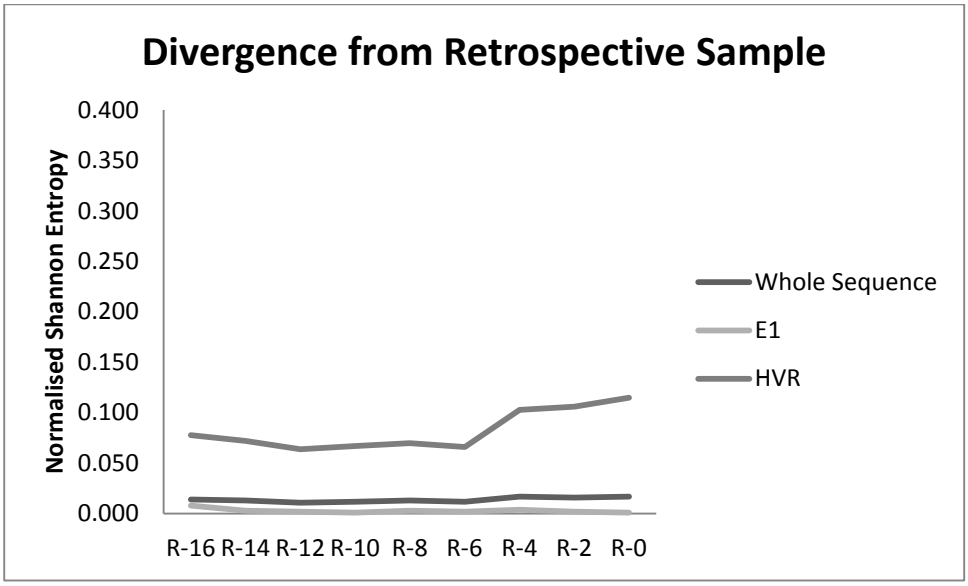
J.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased complexity relative to E1 throughout the study period.



J.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

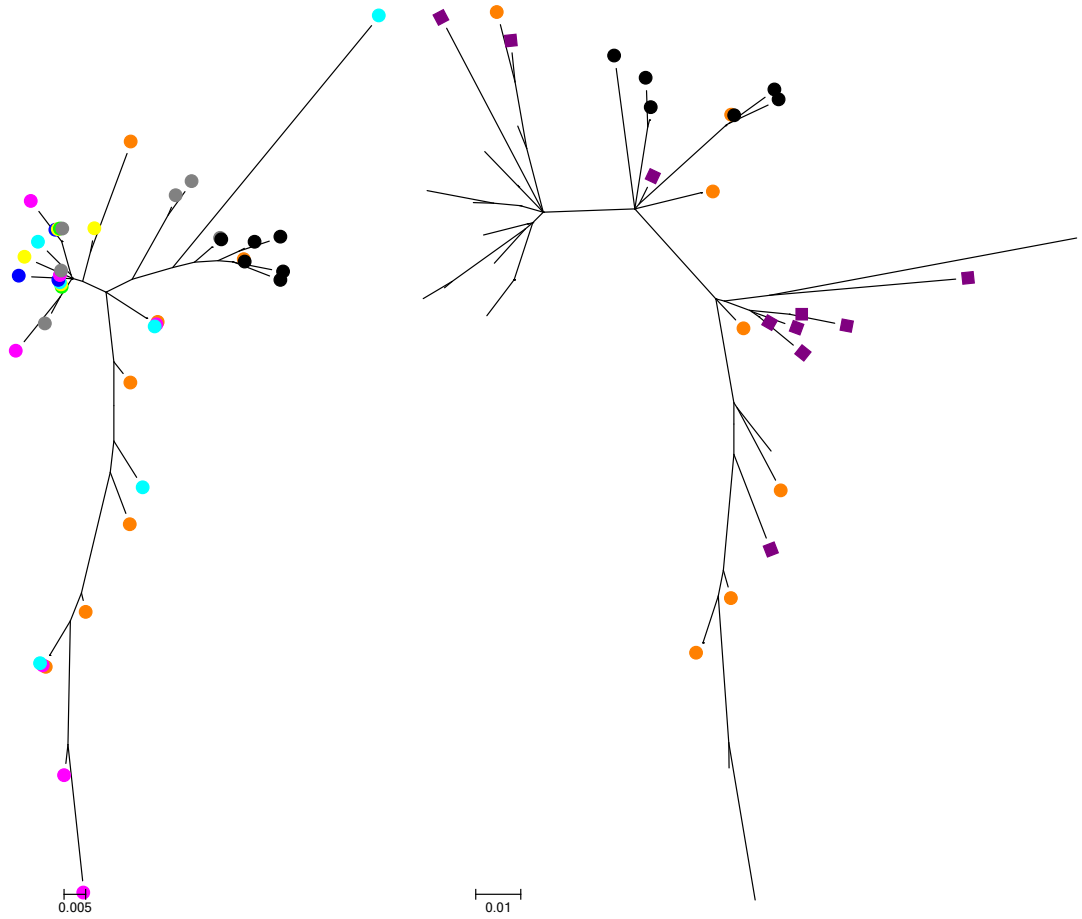
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that there is a similar magnitude of divergence.



J.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group increases throughout the study period and is maximal at the pre treatment sample.

J.1 Phylogenetic analysis



J.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (248 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

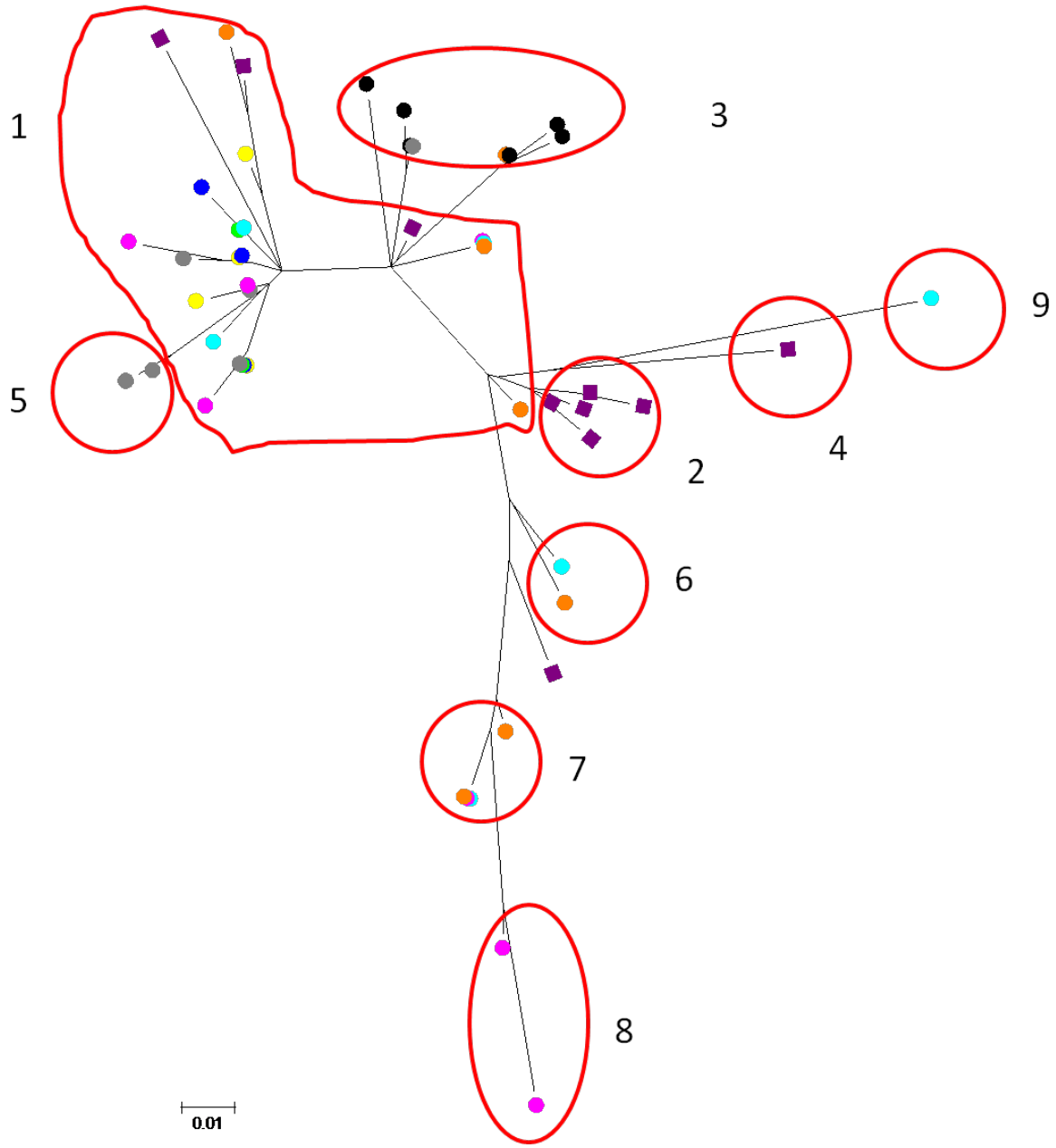
It is noticeable that the general shape of the tree has been affected by the inclusion of the retrospective sample. An outlier turquoise sample has been drawn by the retrospective sample below the main group of sequences at the top of the tree.

J.1 Subpopulation analysis

```
                          1        10         20      27
                          |--------+---------+------|
338010_M13uni_--_10.      TTYTSGGSAAHGAYGIASLFRSGPKQK
d-16-11_M13uni_--_9.      TTYTSGGSAAHGAYGIASLFRSGPKQK
d-16-24_M13uni_--_10      TTYTSGGSAAHGAYGIASLFRSGPKQK
d-16-6_M13uni_--_20.      TTYTSGGSAAHGAYGIASLFRSGPKQK
D-14-11_M13uni_--_20      TTYTSGGSAAHGAYGLASLFSSGPKQK   3
d-16-16_M13uni_--_10      TTYTSGGSAAHGAYGLASLFSSGPKQK
d-16-17_M13uni_--_9.      TTYTSGGSAAHGAYGIASLFSSGPKQK
d-16-19_M13uni_--_58      TTYTSGGSAAHGAYGIASLFSPGPKQK
D-7_M13uni_--_12..89      TTYTSGGSAAHGAYGIASLFSPGPKQK
D-2-15_M13uni_--_20.      TTYTSGGSAAQAAYGLTSLFRLGPKQK
D-2-18_M13uni_--_19.      TTYTSGGSAAHAAYGLTSLFRLGPKQK
3380-4-10_M13uni_--_      TTYTSGGSAAHAAYGLTSLFRLGPKQK
D-6-10_M13uni_--_13.      TTYTSGGSAAHAAYGLTSLFRLGPKQK
3380-8-17_M13uni_--_      TTYTSGGSAAHAAYGLTSLFRLGPKQK
3380-8-3_M13uni_--_7      TTYTSGGSAAHAAYGLTSLFRLGPKQK
D-10-1_M13uni_--_20.      TTYTSGGSAAHAAYGLTSLFRLGPKQK
D-14-16_M13uni_--_14      TTYTSGGSAAHAAYGLTSLFRLGPKQK
D-14-20_M13uni_--_9.      TTYTSGGSAAHAAYGLTSLFRLGPKQK
3380-4-13_M13uni_--_      TTYTSGGSAAHAAYGLTSLFKLGPKQK
3380-4-18_M13uni_--_      TTYTSGGNAAHAAYGLTSLFRLGPKQK
D-6-1_M13uni_--_23..      TTYTSGGNAAHAAYGLTSLFRLGPKQK   1
3380-8-1_M13uni_--_2      TTYTSGGNAAHAAYGLTSLFRLGPKQK
D-10-10_M13uni_--_13      TTYTSGGNAAHAAYGLTSLFRLGPKQK
D-14-18_M13uni_--_10      TTYTSGGNAAHAAYGLTSLFRLGPKQK
D-2-24_M13uni_--_7..      TTYTSGGSAAHTAYGLTSLFRLGPKQK
3380-4-12_M13uni_--_      TTYTSGGSAAHTAYGLTSLFRLGPKQK
3380-0-11_M13uni_--_      TTYTSGGSAAHTAYGLTSLFRLGPKQK
D-22_M13uni_--_24..9      TTYTSGGSAAHGAYGLTSLFRLGPKQK
3380-8-5_M13uni_--_8      TTYTSGGSAAHAVYGLTSLFRLGPKQK
D-6-24_M13uni_--_9..      TTYTSGGSAAHAAYVLTSLFRLGPKQK
3380-0-19_M13uni_--_      TTYTSGGSAAHAAYGLTRLFRLGAKQK
D-11_M13uni_--_11..8      TTYTSGGSAAHAAYGLSSLFSSGPKQK
3380-0-15_M13uni_--_      TTYTSGGSAAHAAYGLSSMFSLGPKQK
D-14-1_M13uni_--_10.      TTYTSGGSVAQGAYGLTSLFRLGPKQK   5
D-14-4_M13uni_--_12.      TTYTSGGSVAQGAYGLTSLFRSGPKQK
D-10_M13uni_--_11..8      TTYTSGGSAAHGAYGITSLFSVGAKQK   2
D-3_M13uni_--_8..895      TTYTSGGSAAHGAYGLTSLFSVGAKQK
D-15_M13uni_--_9..93      TTYTSGGSAAHGAYGLTRLFSVGAKQK
D-5_M13uni_--_7..930      TTYTSGGSAAHGAYGLTRLFSVGAKQK   4
D-12_M13uni_--_11..9      TTYTSGGSAAHGALGLTRLFSVGAKQK
D-16_M13uni_--_29..9      TTYISGGSAAHGVYGLTRLFSPGAKQN   8
D-2-1_M13uni_--_8..9      TTYTSGGSAAHGAYGLASWFSRGAQQK
D-2-16_M13uni_--_18.      TTYTSGGSAAHDALGFTRLFSQGAKQK
3380-0-1_M13uni_--_1      TTYTSGGSAAHDALGFTRLFSQGAKQK   6
3380-4-17_M13uni_--_      TTYTSGGSAAHDALGFTRLFSQGAKQK
   d-16-11_M13uni_--      TTYTSGGSAAHDALGFTRLFRLGAKQK
3380-0-17_M13uni_--_      TTYTSGGSAAHTALGFTRLFSLGPKQK   7
3380-4-15_M13uni_--_      TTYTSGGSAGHDALGLTRLFSLGAQQN
3380-4-16_M13uni_--_      TTYTSGGSAAHDALGFTRLFSLGAQQN   9
D-2-2_M13uni_--_273.      TTYTSGGSAAHTAHGFTRLFRLGAKQK
          Consensus       TTYTSGGSAAH.AyGltsLFslGpkQk
```

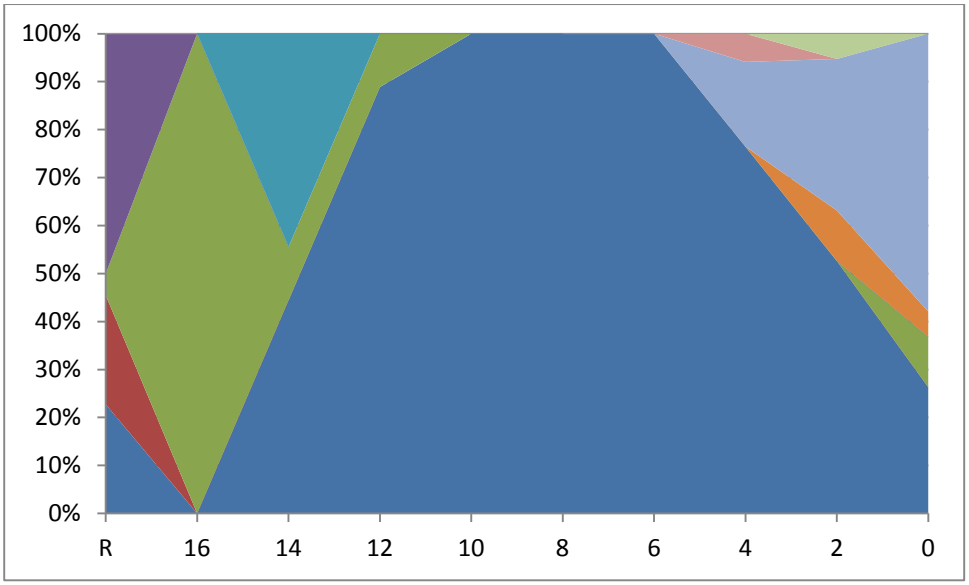
J.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are

designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



J.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

The inclusion of the retrospective samples results in two new subpopulations (2+4).

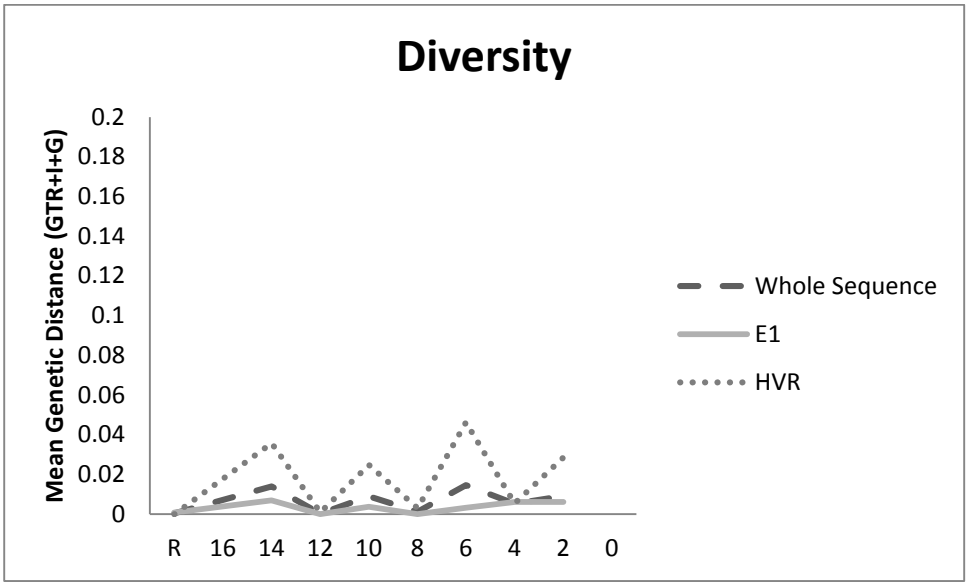


J.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

The most prevalent subpopulation changes 3 times during the study period from subpopulation 4 in the retrospective sample to subpopulations 3 at week 16, 1 at week 14 and finally subpopulation 7 at the pre treatment sample.

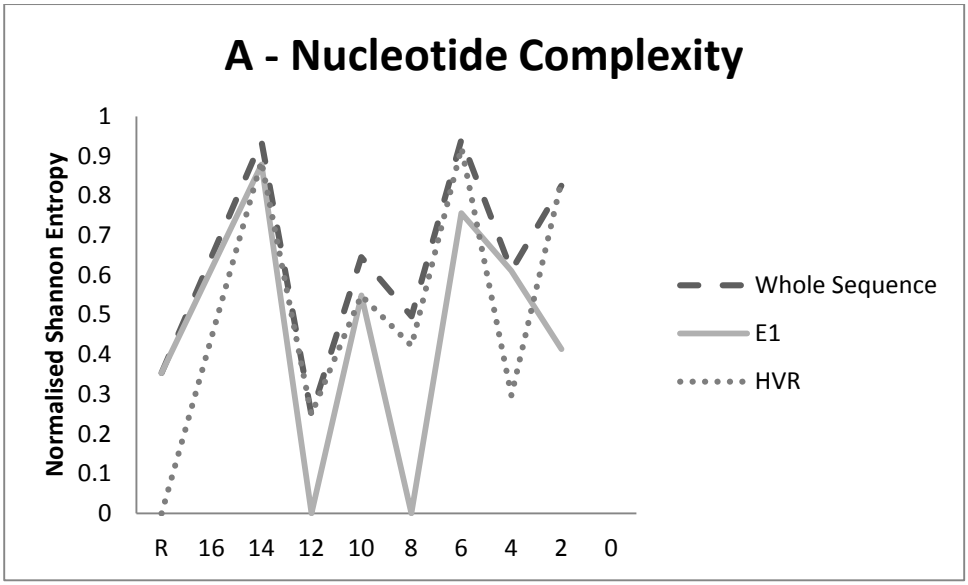
K.1 Subject K

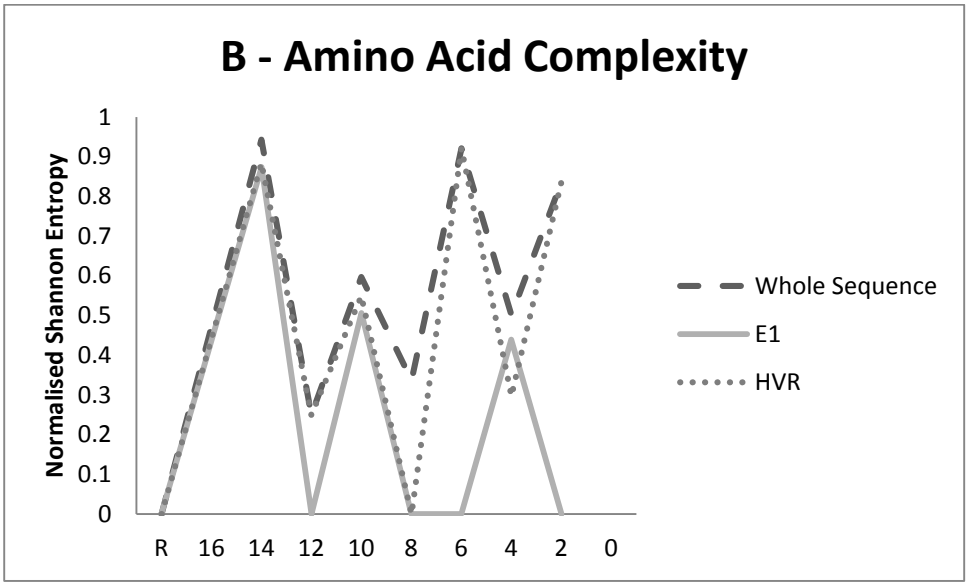
K.1 Diversity, Complexity, and Divergence



K.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

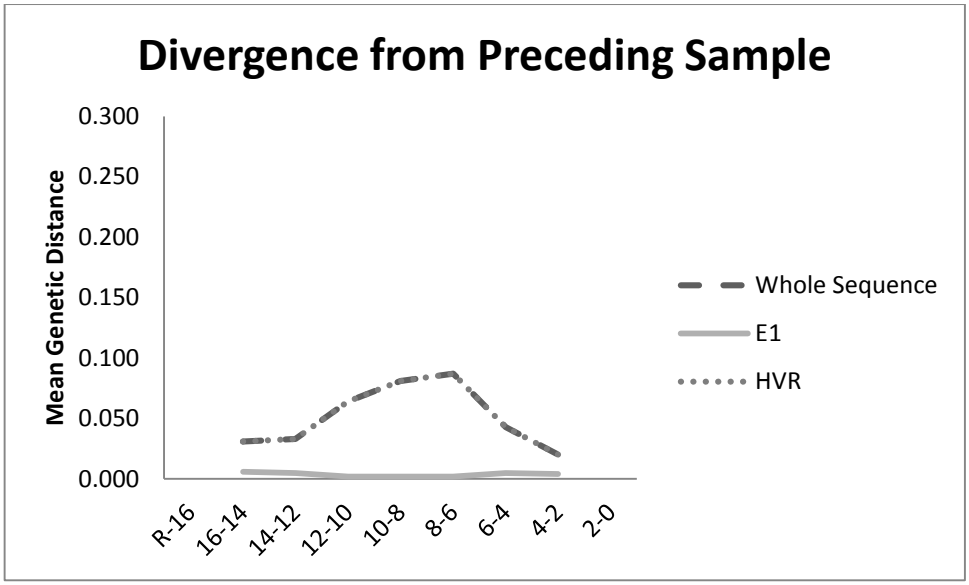
HVR1 diversity is far greater than E1 diversity.





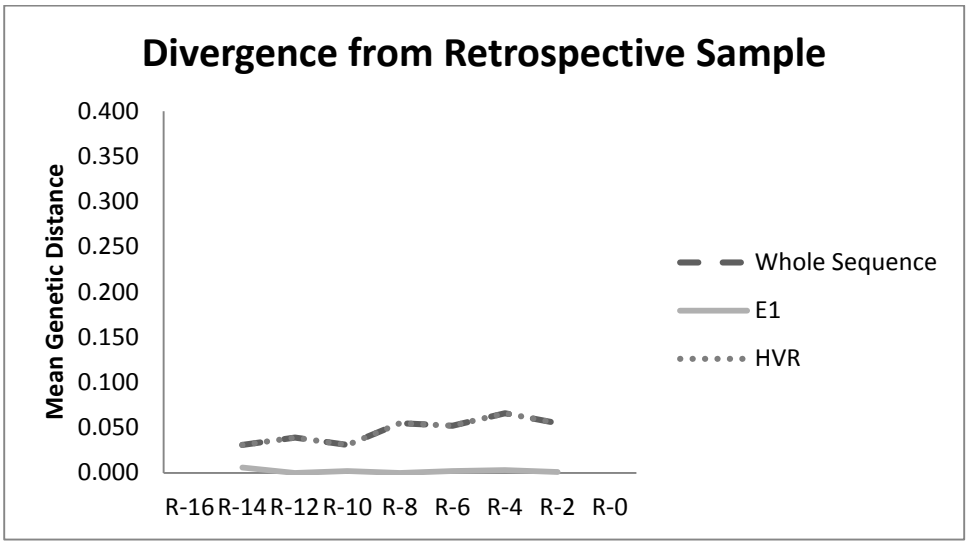
K.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased complexity relative to E1 throughout the study period.



K.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

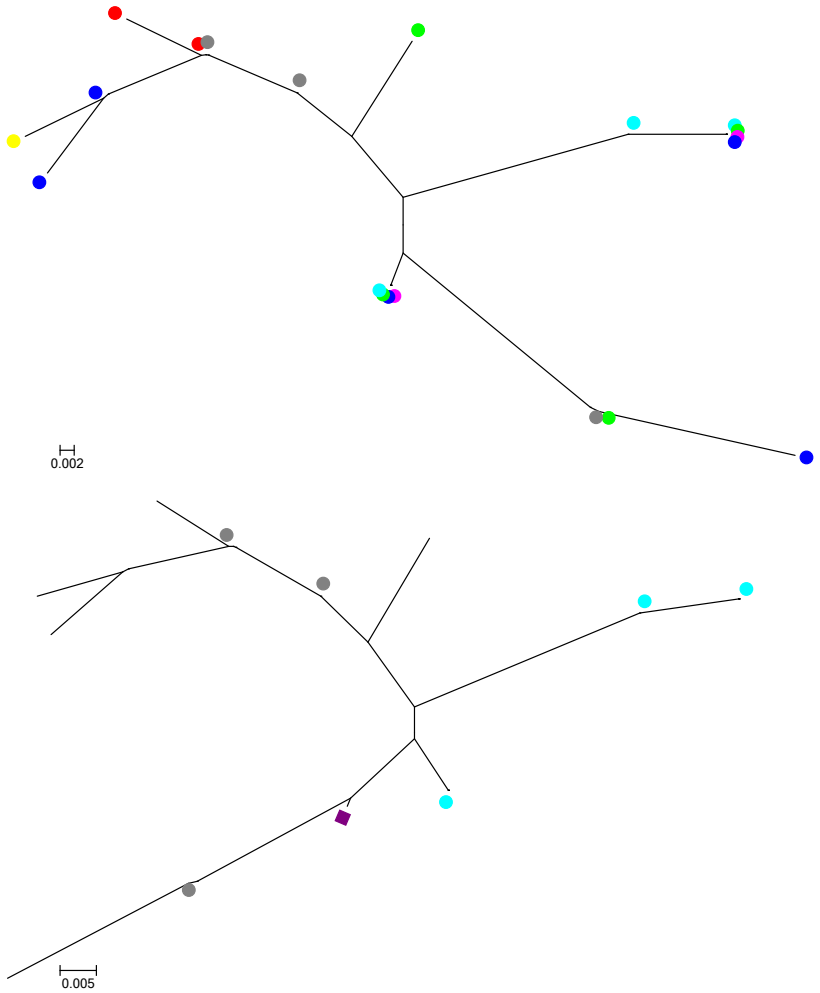
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that the magnitude of divergence is greater for each fortnight for the three intervals between weeks 12 and 6.



K.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group increases throughout the study period and is maximal at the week 2 sample.

K.1 Phylogenetic analysis



K.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (314 days prior to Week 14 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise. Identical sequences overlap.

It is noticeable that the general shape of the tree is unaffected by the inclusion of the retrospective sample.

K.1 Subpopulation analysis

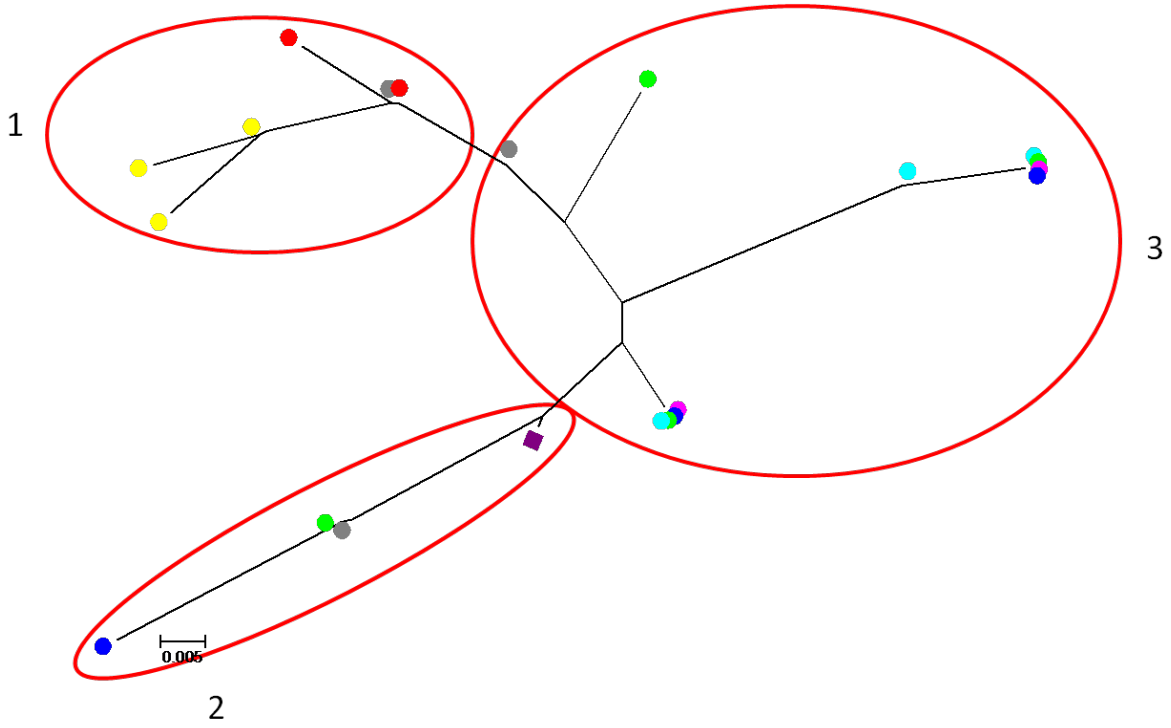
```
                        1        10         20      27
                        |--------+---------+------|
0105-2-10_M13uni_--_    GTRISGGSAAYATSRFTGLFSSGPSQK
E4-2_M13uni-21_--_42    GTRISGGSAAYATSRFTGLFSSGPSQK
0105-6-12_M13uni_--_    GTRISGGSAAYATSRFTGLFSSGPSQK
E-10-1_M13uni_--_23.    GTRISGGSAAYATSRFTGLFSSGPSQK
0105-2-9_M13uni_--_1    GTRISGGSAAHATSRFTGLFSSGPSQK
E4-16_M13uni-21_--_2    GTRISGGSAAHTVSRFTGLFSSGPSQK
0105-6-1_M13uni_--_1    GTRISGGSAAHTVSRFTGLFSSGPSQK
0105-2-19_M13uni_--_    GTRISGGSAAHTVSRFTGLFSSGPSQK
E-10-20_M13uni_--_10    GTRISGGSAAHTVSRFTGLFSSGPSQK
E-10-12_M13uni_--_24    GTRISGGSAAHTINRFTGLFSSGPSQK
E14-12_M13uni-21_--_    GTHISGGSAAHTINRFTGLFSSGPSQK
0105-6-14_M13uni_--_    ETYISGGSAAHTVSRFTGLFSSGPSQK
E-10-10_M13uni_--_9.    ETYISGGSAAHTVSRFTGLFSSGPSQK
E14-9_M13uni-21_--_9    ETYISGGSAAHTVSRFTGLFSSGPSQK
E-A-10_M13uni_--_45.    GTYISGGSAAHTISRFTGLFSSGPSQK
0105-8-9_M13uni_--_2    GTHISGGSAARTINRFTGLFSSGPSQK
0105-8-16_M13uni_--_    GTHISGGSAARTINRFTGLFSSGPSQK
0105-8-18_M13uni_--_    GTHISGGSAARTINRFTGLFSSGPSQK
E-12-1_M13uni_--_20.    GTHISGGSAARTINRFTGLFSSGPSQK
E14-3_M13uni-21_--_7    GTHISGGSAARTINRFTGLFSSGPSQK
E-12-18_M13uni_--_10    GTHISGGSAARTINRFTGLFSSDPSQK
           Consensus    gTrISGGSAAht.sRFTGLFSSGPSQK
```

3

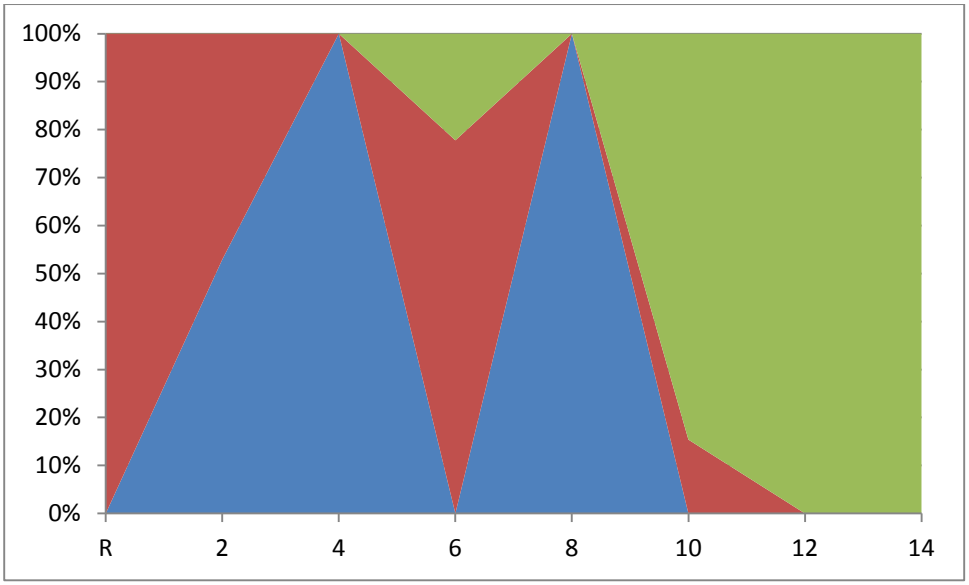
2

1

K.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



K.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise. Identical sequences overlap.

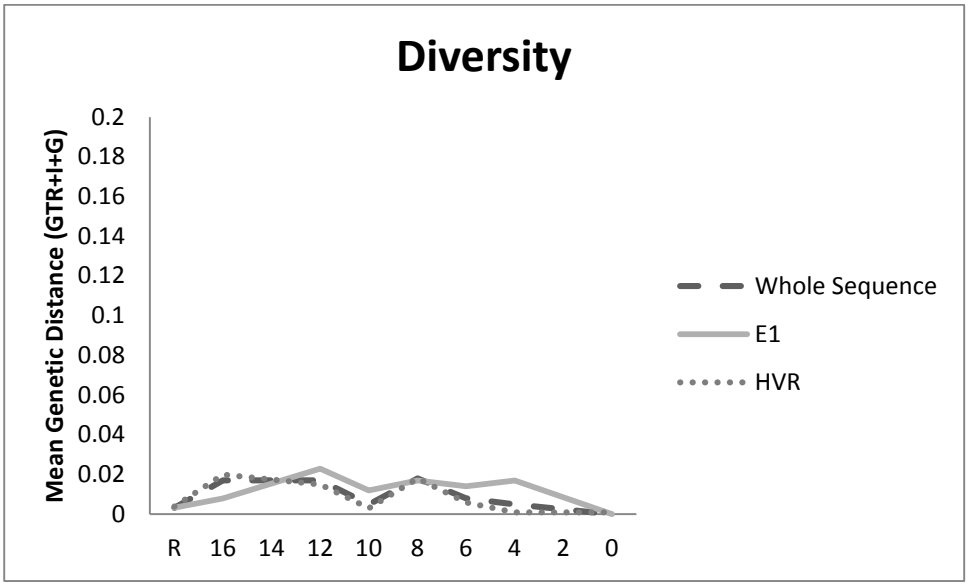


K.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

The most dominant subpopulation at the beginning of the study had sustained this dominance since the retrospective sample 10 months previously. A new subpopulation appears at week 14 and by the pre treatment sample this has completely replaced the other two subpopulations by the pre treatment sample. Examination of the sequence alignment however suggests that these subpopulations are relatively closely related to each other with not all sequences within each subpopulation differing from the other subpopulations by more than three amino acids. This suggests a broad flat fitness optimum which is being thoroughly investigated by the virus QS.

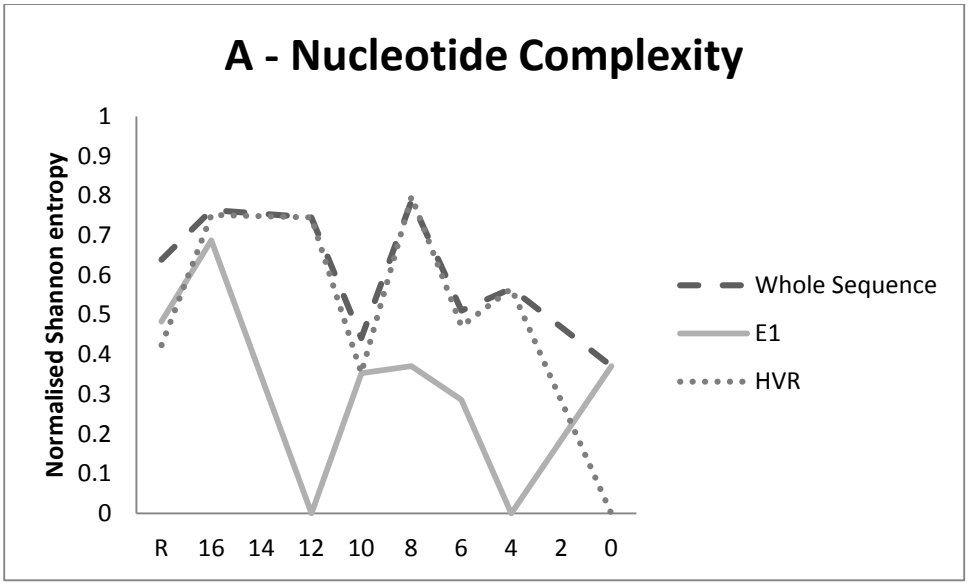
M.1 Subject M

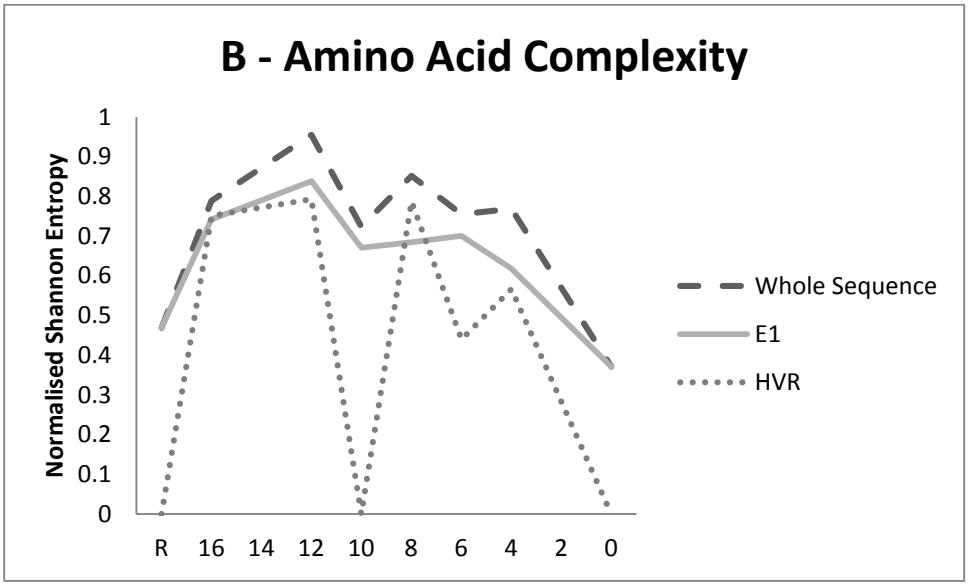
M.1 Diversity, Complexity, and Divergence



M.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

HVR1 diversity is similar to E1 diversity.

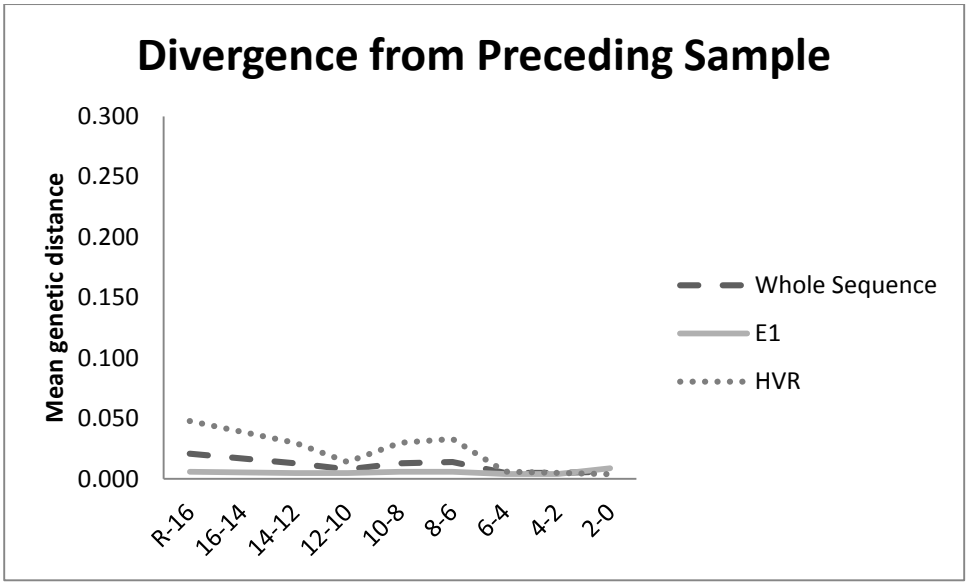




M.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

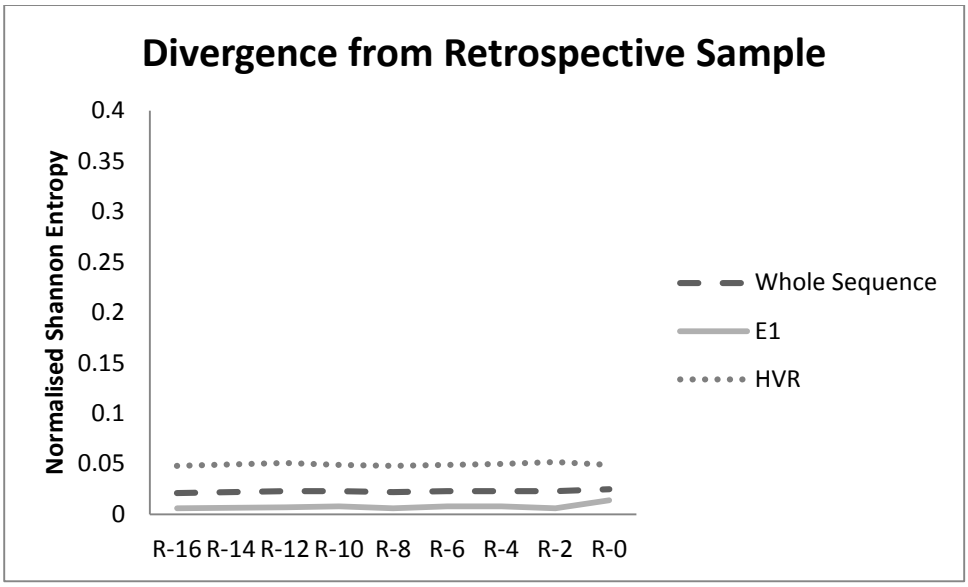
HVR1 demonstrates increased complexity relative to E1 at many points during the study period.

Subject M is unusual however in demonstrating a homogenous single HVR1 amino acid profile at three times during the study period which characterised by a complexity of 0. This potentially suggests multiple episodes of purifying selection.



M.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

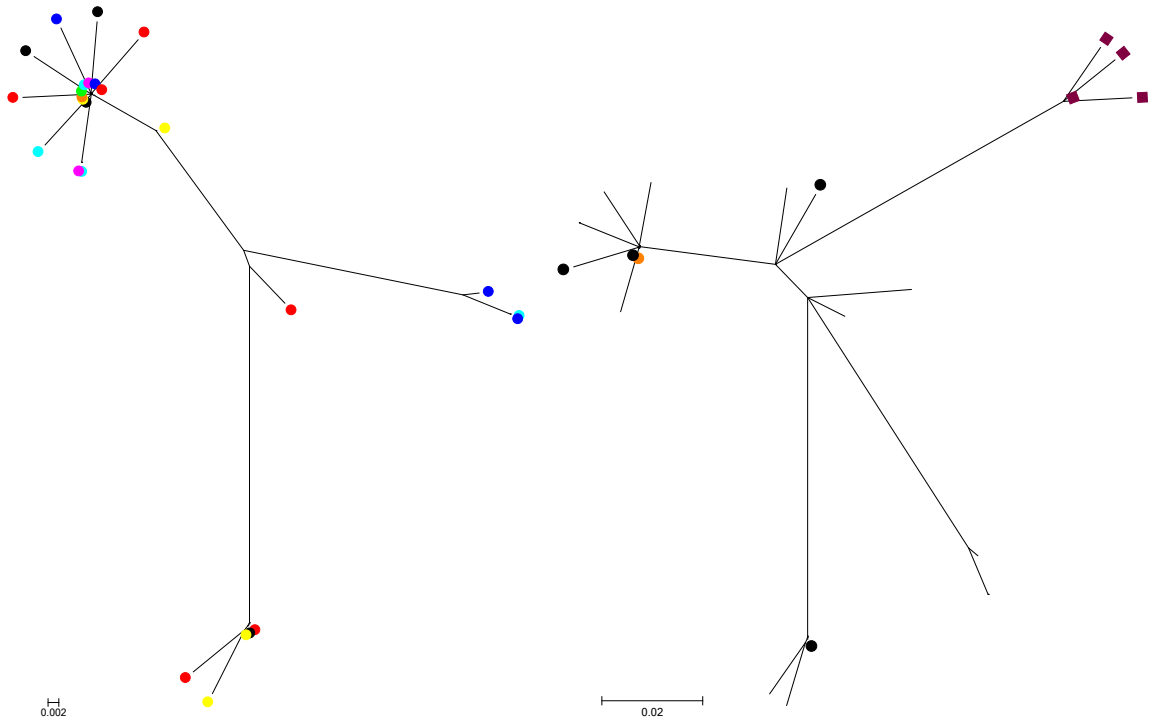
E1 demonstrates almost no divergence throughout the study period. HVR1 appears to diverge between the retrospective sample and the sample 16 weeks prior to commencing treatment but thereafter there is minimal change in the HVR1 profile.



M.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group is of almost the identical magnitude between it and all subsequent samples.

M.1 Phylogenetic analysis



M.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (287 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree is significantly changed by the inclusion of the retrospective sample. The retrospective clones all group in a clade remote from the study sequences.

This suggests that following the elimination of the retrospective clade that the HVR1 eplored the sequences space in a multi directional fashion but that by the pre treatment sample, a new dominant clade has become established (orange marker).

M.1 Subpopulation analysis

```
                        1       10        20     27
                        |--------+---------+------|
15090450905509065090   ATYITGGASAHTTYGLASLFTAGPRQK
 G2-1_M13uni-21_--_2    ATYITGGASAHTTYGLASLFTAGPRQK
G2-8_M13uni-21_--_50    ATYITGGASAHTTYGLASLFTAGPRQK
150943_M13uni50944_M    ATYITGGASAHTTYGLASLFTAGPRQK
1509-6-1_M13uni_--_9    ATYITGGASAHTTYGLASLFTAGPRQK
150983_M13uni50987_M    ATYITGGASAHTTYGLASLFTAGPRQK
1509-101_M13uni_--_8    ATYITGGASAHTTYGLASLFTAGPRQK
      1509124_M13uni    ATYITGGASAHTTYGLASLFTAGPRQK
1509127_M13uni509121    ATYITGGASAHTTYGLASLFTAGPRQK
150981612_M13uni5091    ATYITGGASAHTTYGLASLFTAGPRQK
    150916841_M13uni    ATYITGGASAHTTYGIASLFTAGPRQK
     15091215_M13uni    ATYITGGASAHTTYGLTSLFTAGPRQK
      1509817_M13uni    TTYITGGASAHTTYGLASLFTAGPRQK
G-2-192-16_M13uni-21    ATYITGGASAHTAYGLASLFTAGPRQK
150942_M13uni50947_M    ATYITGGASAHTAYGLASLFTAGPRQK
1509-6-16_M13uni_--_    ATYITGGASAHTTYGLASLFNAGPRQK
    150916810_M13uni    ATYITGGASAHATYGLASLFTAGPRQK
G2-10_M13uni-21_--_1    TTYITGGASAHTTFGLASLFTKGAQQK
1509-6-13_M13uni_--_    TTYITGGASAHTTFGLASLFTKGAQQK
1509-6-20_M13uni_--_    TTYITGGASAHTTFGLASLFTKGARQK
      1509122_M13uni    TTYISGGASAHTTYGLASLFTAGAQQK
150981_M13uni50982_M    TTRITGGASAHNTYGLASLFSVGAQQK
       150984_M13uni    TTRITGGASAHNTYGLASLFSVGAQQK
     15091214_M13uni    TTRITGGASAHNTYGLASLFSVGAQQK
15091687_M13uni50916    TTRITGGASAHNTYGLASLFSVGAQQK
      1509121_M13uni    TTRISGGASAHNTYGLASLFSVGAQQK
G-10_M13uni-21_--_9.    TTHITGGASAHATYGLASLFKVGARQK
G-17_M13uni-21_--_10    TTHITGGASAHATYGLASLFKVGARQK
G-19_M13uni-21_--_7.    TTHITGGASAHATYGLASLFKVGARQK
G-6_M13uni-21_--_11.    TTHITGGASAHATYGLASLFKVGARQK
           Consensus    tTyItGGASAHtT%GLASLFt.GarQK
```

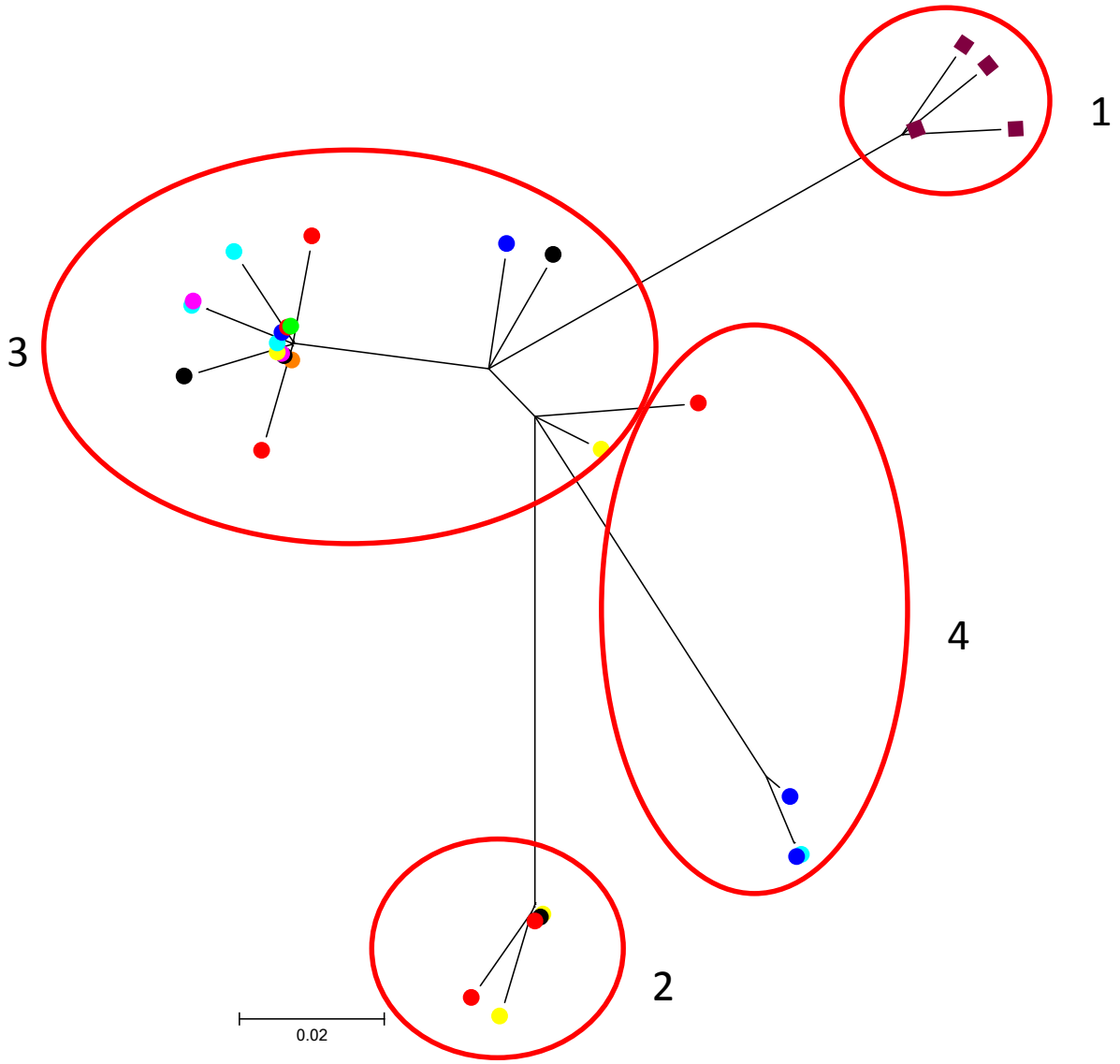
3

4

2

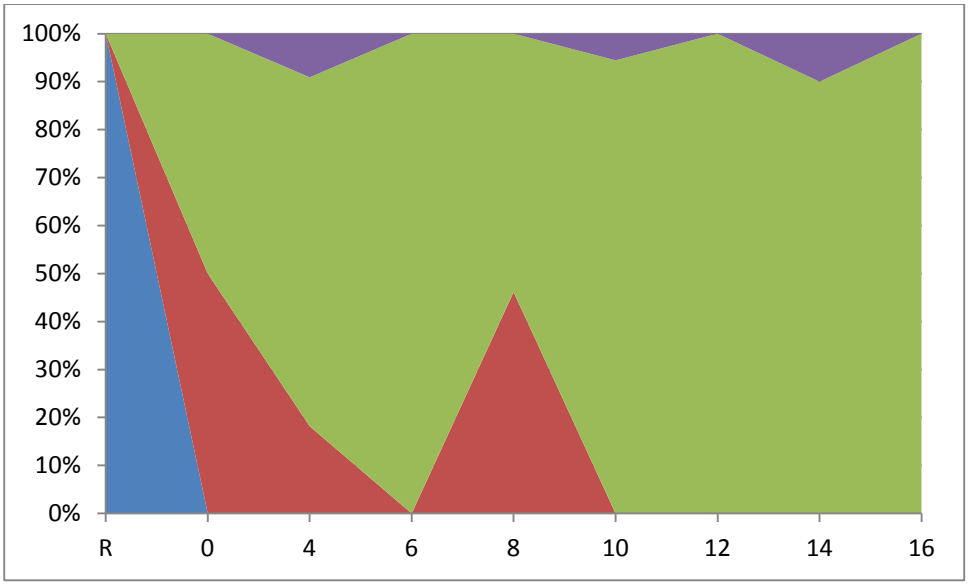
1

M.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



M.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

This illustrates the description of a new subpopulation with the inclusion of the retrospective sample.

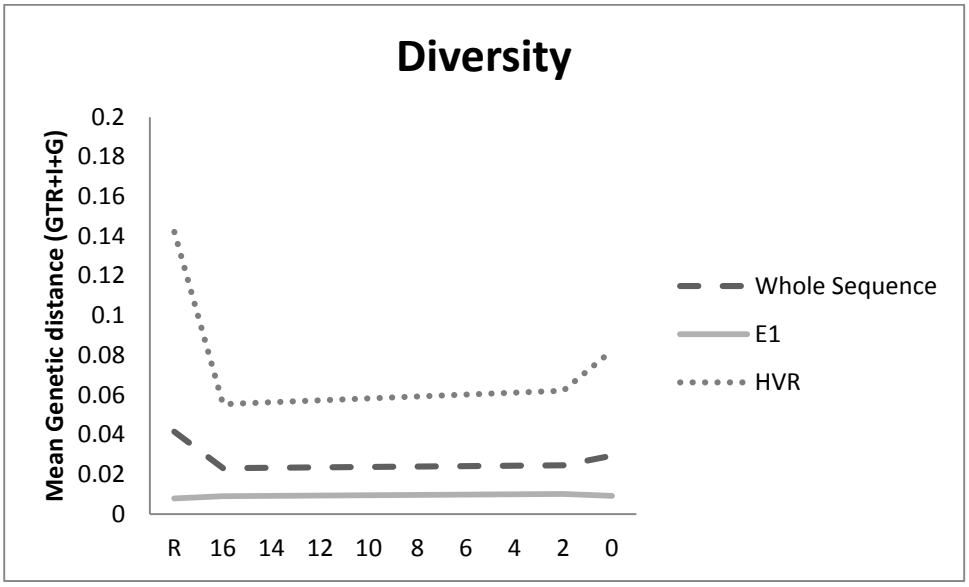


M.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

The initial dominant subpopulation is no longer present at the week 0 sample which is taken 16 weeks before treatment is commenced. At week 0 there is co dominance between two new subpopulations but as the study progresses, subpopulation 3 comes to dominate the entire milleu ar has been suggested by Fig. 5.

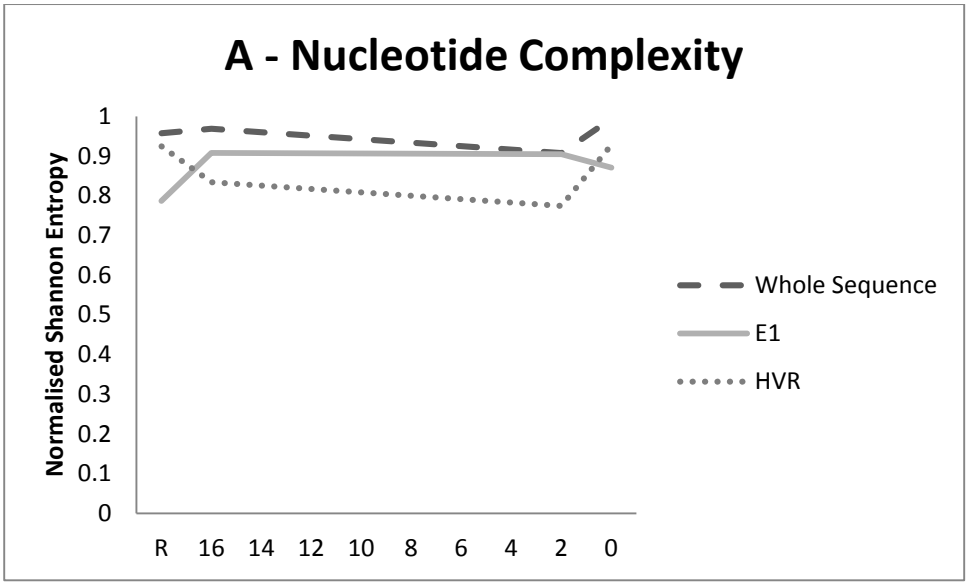
N.1 Subject N

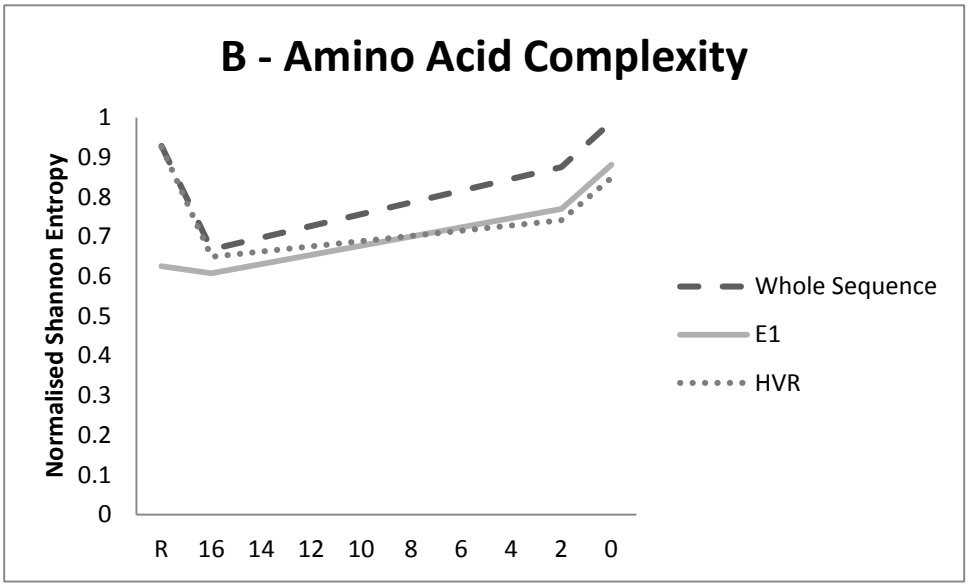
N.1 Diversity, Complexity, and Divergence



N.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

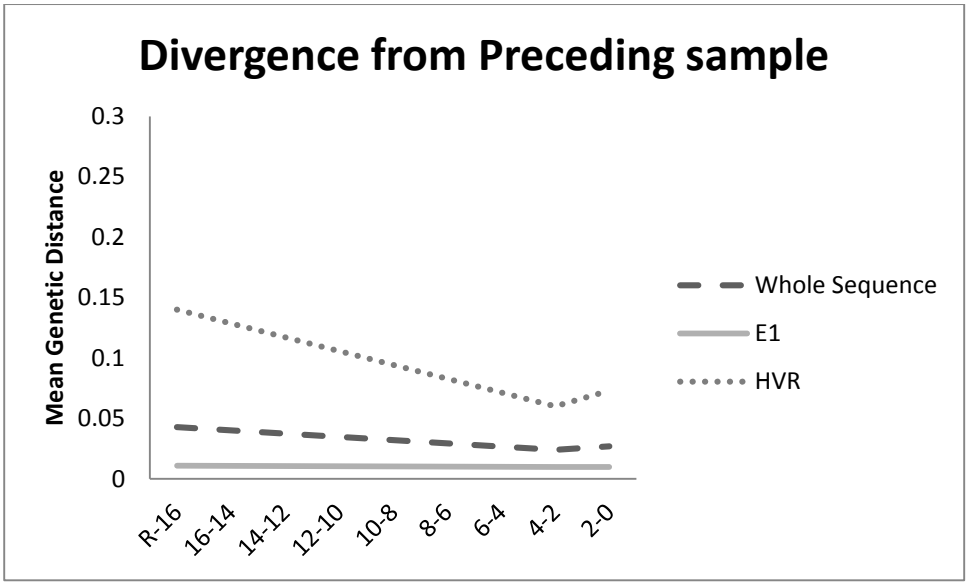
HVR1 diversity is greater than E1 diversity.





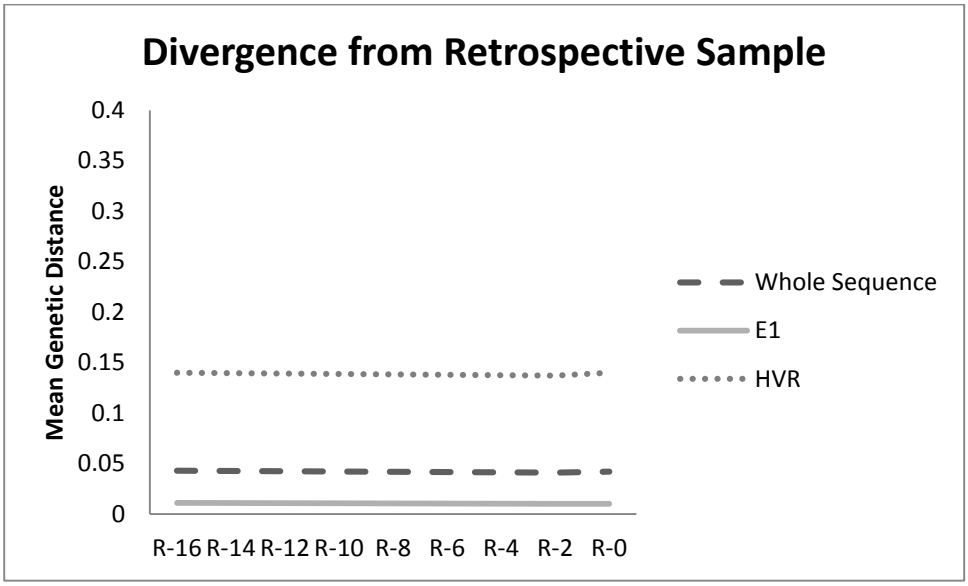
N.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates similar complexity when compared with E1 throughout the study period and it is notable that E1 complexity is more than HVR1 complexity both at nucleotide and amino acid level in half of the samples. This suggests little positive or purifying selection.



N.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

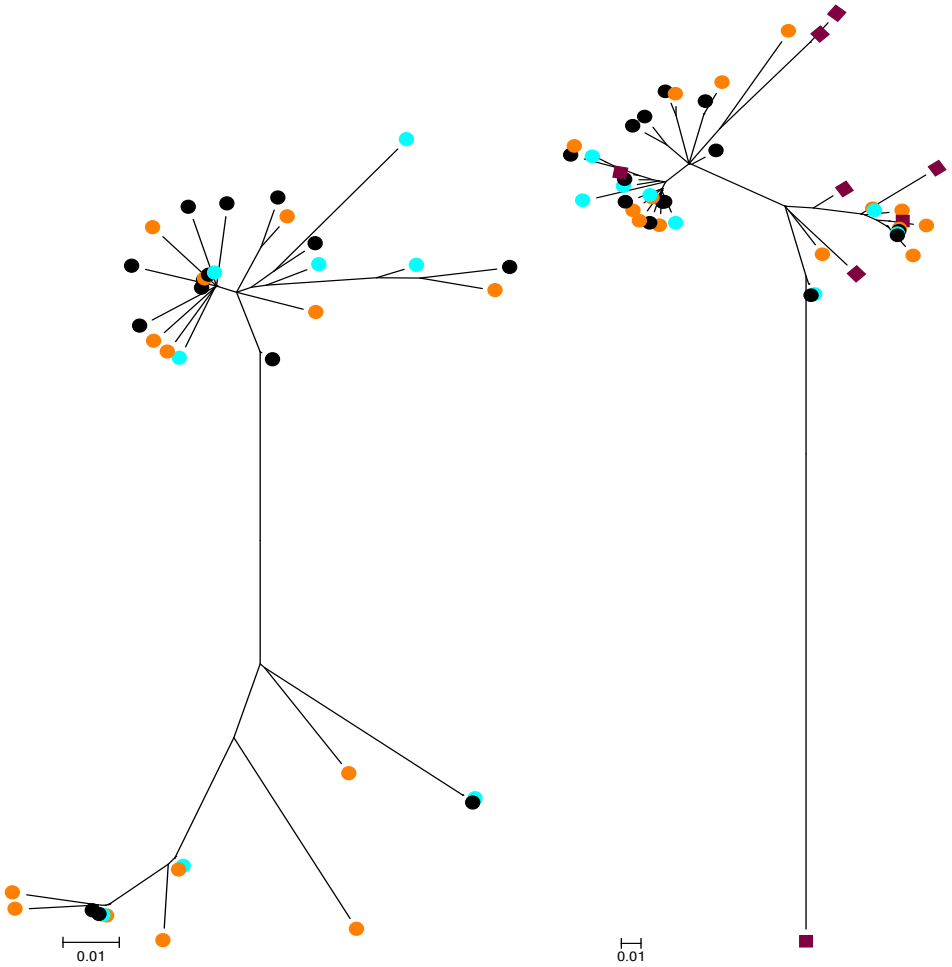
HVR1 divergence between the retrospective sample and week 16 pre treatment is 0.14 potentially suggesting significant sequence change.



N.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group remains constant for each of the subsequent study samples.

N.1 Phylogenetic analysis



N.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (465 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree is changed by the inclusion of the retrospective sample. The change in the scale distance highlights that the componenents of Fig 2 now form the collection of sequences seen together at the top of Fig 3. One of the retrospective sequences is clearly remotely related to the remainder of the HVR1 QS. This could explain the divergence data with a significant proportion of the pairwise divergence by the inclusion of this remote sequence. Furthermore, this also explains the very high degree of diversity described for HVR1 in the retrospective sample (Table 3).

N.1 Subpopulation analysis

```
                        1        10         20      27
                        |--------+---------+------|
2308-0308-0-91_M13un    ETHVTGGTAAHGALGITSLLSRGPKQN
2308-2-10_M13uni_--_    ETHVTGGTAAHGALGITSLLSRGPKQN
2308-2-21_M13uni_--_    ETHVTGGTAAHGALGITSLLSRGPKQN
62_M13uni7_M13uni5_M    ETHVTGGTAAHGALGITSLLSRGPKQN
            68_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            69_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            61_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
    2308-0-85_M13uni    ETHVTGGTAAHGALGITSLLSRGPKQN
            64_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
    2308-0-88_M13uni    ETHVTGGTAAHGALGITSLLSRGPKQN
79_M13uni8_M13uni6_M    ETHVTGGTAAHGALGITSLLSRGPKQN
            74_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
            80_M13uni   ETHVTGGTAAHGALGITSLLSRGPKQN
2308-2-1_M13uni_--_1    ETHVTGGTAVHGALGITSLLSRGPKQN
    2308-0-86_M13uni    ETHVTGGTVAHGALGITSLLSRGPKQN
    2308-0-90_M13uni    ETHVTGGTAAHGALGITSLLSRGPNQN
            77_M13uni   ETHVTGGTTAHGAFGITSLLSRGPKQN
    2308-0-62_M13uni    ETHVTGGTTAHGALGITSLLSRGPKQN
2308-2-16_M13uni_--_    ETHVTGGTAAHGVFGITSLLSRGPKQN
            70_M13uni   ETHVTGGTAAHGVFGITSLLSSGPKQN
            71_M13uni   ETHVTGGTAAHGALGITSLLSPGPKQN
HB7_M13uni-21_--_39.    ETHVTGGTAAHGAFGITSLLSPGPKQN
    2308-0-83_M13uni    ETHVTGGTAAHGVFGITSLLSPGPKQN
2308-2-6_M13uni_--_8    ETHVTGGTAAHGALGISRLLSLGPKQN
2308-2-19_M13uni_--_    DTRVTGGTVAHGALTLTSLLSRGAKQN
            75_M13uni   DTRVTGGTVAHGALTLTSLLSRGAKQN
HB13_M13uni-21_--_17    DTRVTGGTAAHGAYTLTTLLSRGAKQN
HB8_M13uni-21_--_307    ETRVTGGTAAHGALTLTTLLSRGPKQN
2308-2-20_M13uni_--_    ETHVTGGTAAHGAFTLTTLLSRGAKQN
    2308-0-84_M13uni    ETHVTGGTAAHGAFTLTTLLSRGAKQN
    2308-0-87_M13uni    ETHVTGGTAAHGAFTLTTLLSRGAKQN
HB18_M13uni-21_--_96    ETHVTGGTAAHGAFTLTTLLSRGAKQN
2308-2-23_M13uni_--_    ETHVTGGTAAHGAFTLTTLLNRGAKQN
   63_M13uni6_M13uni    ETHVTGGTAAHGAFTLTTLLNRGAKQN
            73_M13uni   ETHVTGGTAAHGAFTLTTLLNRGAKQN
    2308-0-61_M13uni    ETHVTGGTAAHGAFTLTTLLNRGAKQN
    2308-0-63_M13uni    ETHVTGGTAAHGAFTLTTLLNRGAKQN
2308-0-94_M13uni308-    ETHVTGGTAAHGAFTLTTLLNRGAKQN
    2308-0-96_M13uni    ETHVTGGTAAHGAFTLTTLLSRGPKQN
HB14_M13uni-21_--_24    ETHVTGGTAAHGAFTLTNLLSRGAQQN
    2308-0-81_M13uni    ETHVTGGTAAHGALTLTSLLSRGAKQN
HB10_M13uni-21_--_35    ETHVTGGTAAHGALTLSRFLSASPKQN
HB19_M13uni-21_--_9.    ETHVTGGTAAHGALTLSRFLSAGPKQN
HB23_M13uni-21_--_24    YTRVTGGSQAQQARTLANIFSLGPRQN
            Consensus   eThVTGGtaAhgAltlt.llSrGpkQN
```

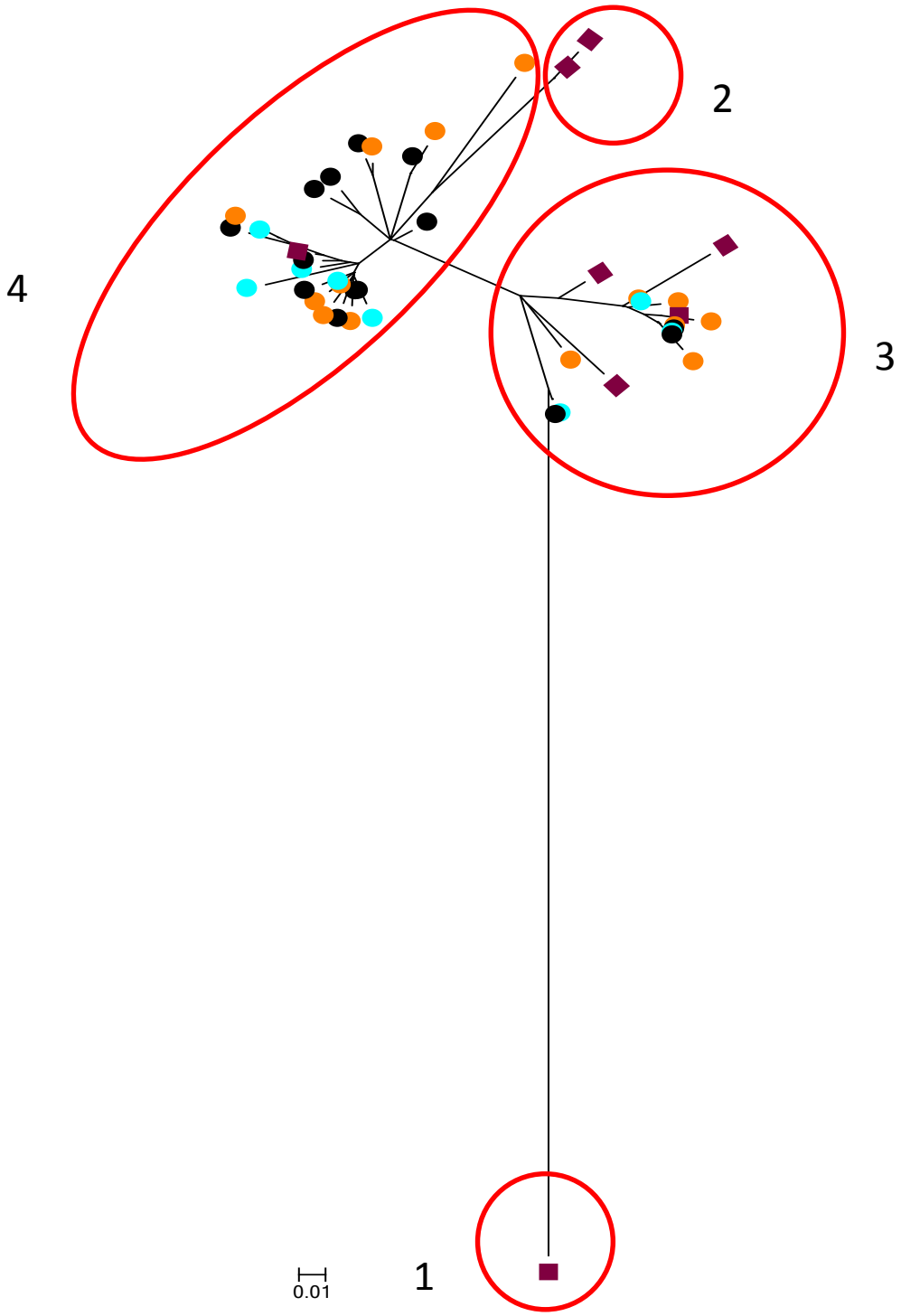
4

3

2

1

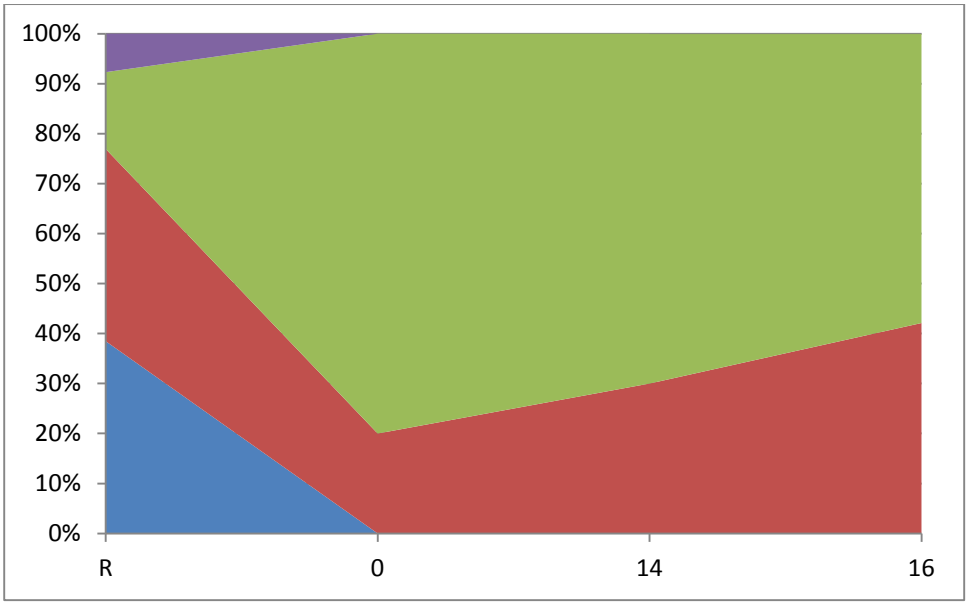
N.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



N.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

The inclusion of the retrospective samples increases the number of subpopulations seen in subject N

to 4.

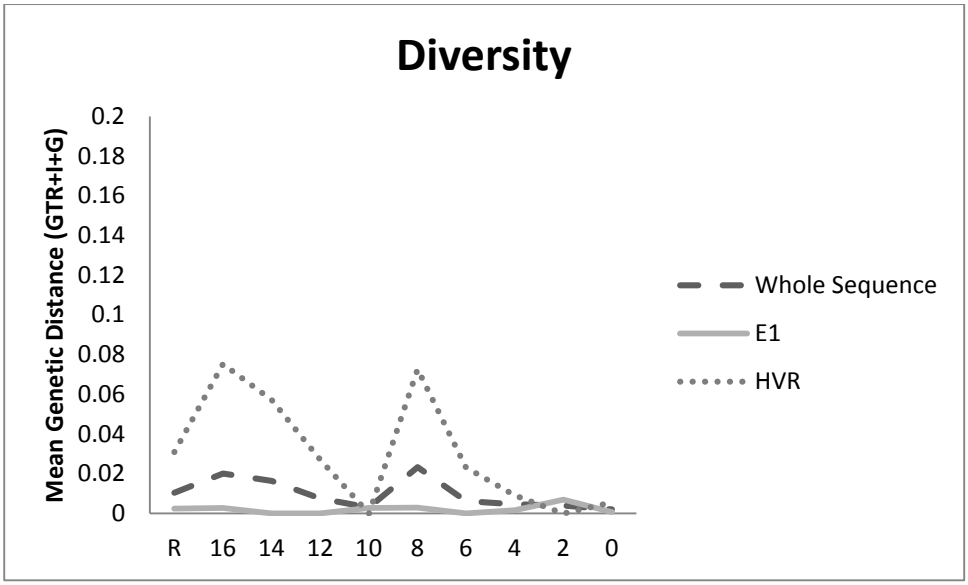


N.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subpopulations 1 + 4 are only present in the retrospective sample. Subpopulations 2 + 3 vie for co dominance throughout the study including the retrospective sample with neither being eliminated at any time. This suggests some initial sequence selection but no further selection during the 16 weeks prior to treatment commencement. It also explains both the high diversity in the retrospective sample and the high degree of divergence between the retrospective and week 16 sample.

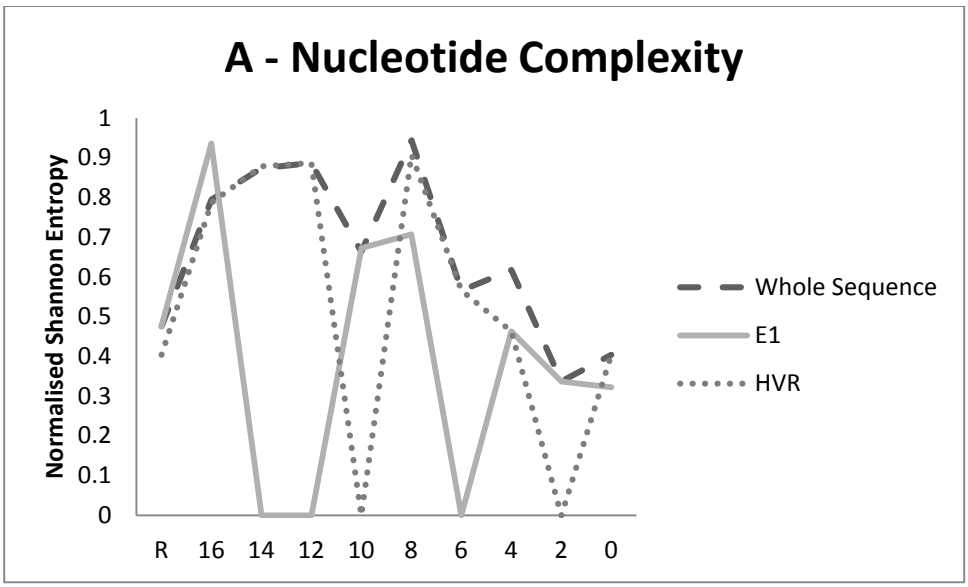
O.1 Subject O

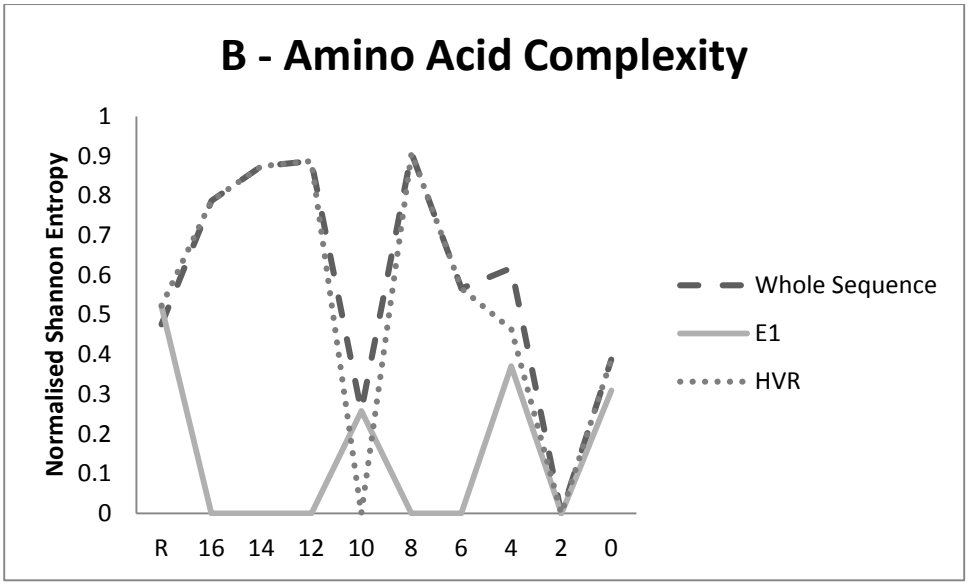
O.1 Diversity, Complexity, and Divergence



O.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

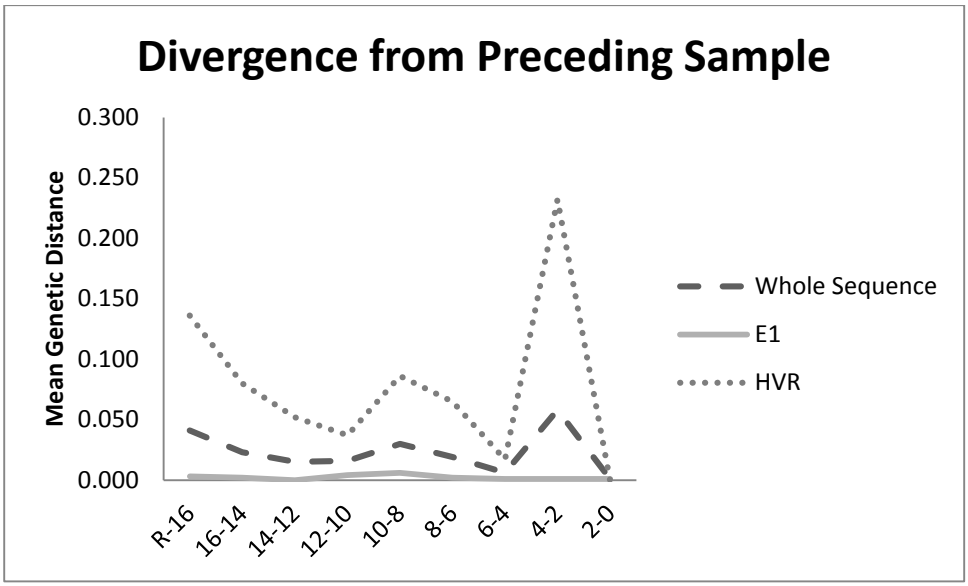
HVR1 diversity is greater than E1 diversity.





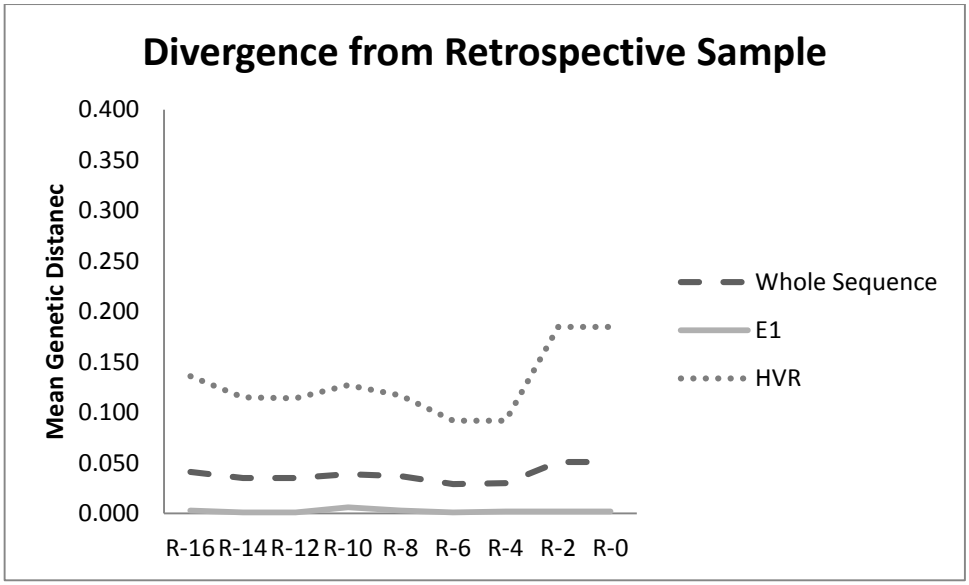
O.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates markedly increased complexity relative to E1 throughout the study period.



O.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

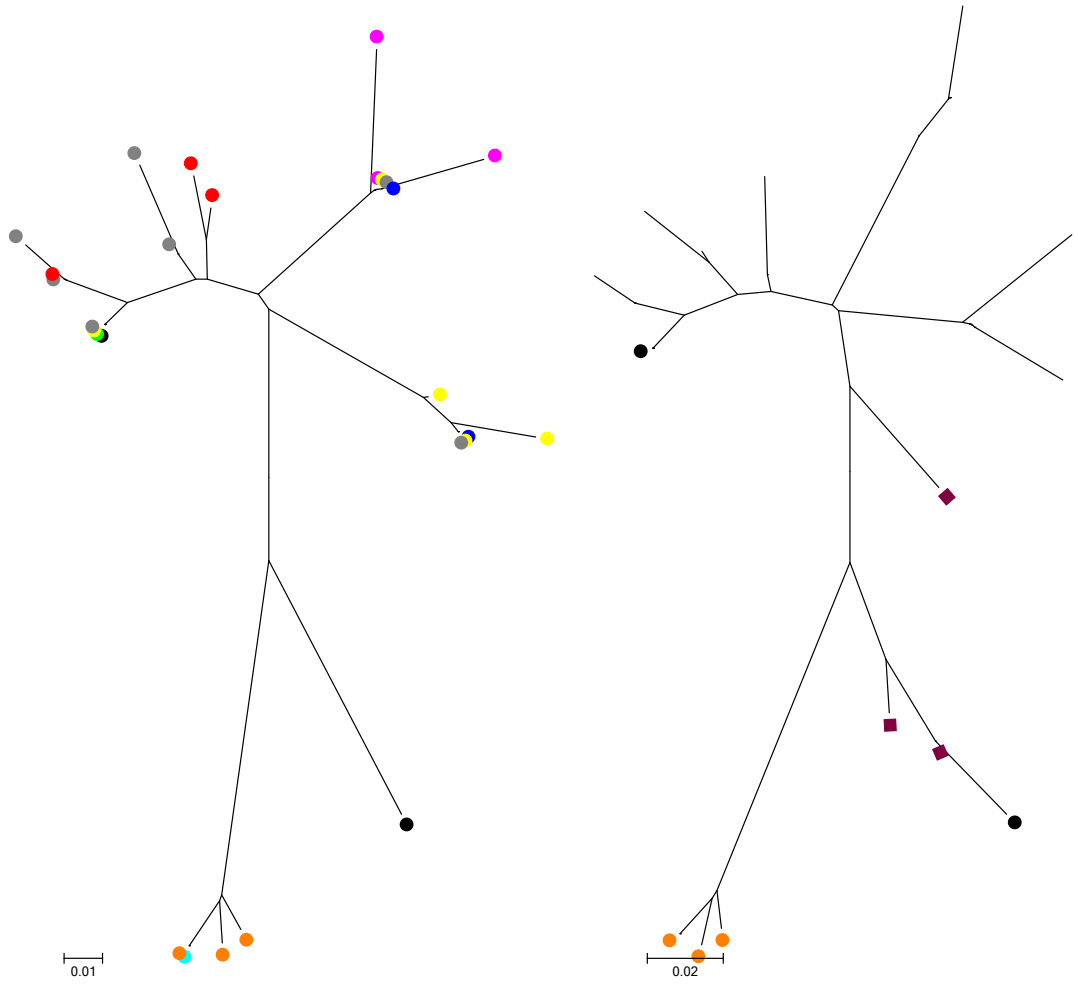
Divergence is greatest between the retrospective sample and the week 16 sample.



O.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group is maximal at the pre treatment sample.

O.1 Phylogenetic analysis



O.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (175 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree is unaffected by the inclusion of the retrospective sample.

O.1 Subpopulation analysis

```
                        1        10         20      27
                        |--------+---------+------|
1812-2-1_M13uni_--_1    STHTIGGSAASNAFVLSHLLSRGPRQN
181289010_M13uni_--_    STHTIGGSAASNAFVLSHLLSRGPRQN
181289014_M13uni_--_    STHTIGGSAASNAFMLGHLLSRGPRQN
18128901_M13uni_--_1    STHTIGGSAASNAFVLGHLLSPGPRQN
  1812-169_M13uni_--    STHTVGGSVARNAFKFTSLLSRGPQQN
I-9_M13uni-21_--_12.    STHTVGGSVARNAFKITSLLSRGPQQN
I-18_M13uni-21_--_18    STHTVGGSVARNAFKITSLLSRGPSQN
1812-161_M13uni_--_1    STHTIGGSMAHSASGLASLFSQGPKQN
1812-8-16_M13uni_--_    STHTIGGSMAHSASGLASLFSQGPKQN
I-10-1_M13uni_--_11.    STHTIGGSMAHSASGLASLFSQGPKQN
1812-1410_M13uni_--_    STHTIGGSMAHSASGLASLFSQGPKQN
1812-1210_M13uni_--_    STRTIGGSMAHSASGIASLFSQGPKQN
1812-1414_M13uni_--_    STRTIGGSMAHSASGIASLFSQGPKQN
1812-148_M13uni_--_1    STRTIGGSMAHSASGIASLFSQGPKQN
1812-141_M13uni_--_1    STHTIGGSMAHSASGIASLFSRGPNQN
1812-1412_M13uni_--_    STHTIGGSMAYSTSGIASLFSRGPNQN
1812-1212_M13uni_--_    STHTIGGSMARRASGIASLFSQGPSQN
1812-122_M13uni_--_1    STHTIGGSMARRASGIASLFSQGPKQN
I-1_M13uni-21_--_21.    STHTIGGTVARGASGIASLLSRGPSQN
1812-4-11_M13uni_--_    STHTIGGSVARSASGIASWFSRGPNQN
1812-6-11_M13uni_--_    STHTIGGSVARSASGIASWFSRGPNQN
1812-8-14_M13uni_--_    STHTIGGSVARSASGIASWFSRGPNQN
1812-1411_M13uni_--_    STHTIGGSVARSASGIASWFSRGPNQN
1812-4-2_M13uni_--_3    TTHTIGGSVARSASGIASWFSRGPNQN
1812-4-5_M13uni_--_2    STHTIRGSVARTASGIASWFSLGPNQN
1812-6-1_M13uni_--_1    STHTIGGSVALRASRIAGLFSRGPNQN
1812-8-12_M13uni_--_    STHTIGGSVALRASRIAGLFSRGPNQN
1812-144_M13uni_--_2    STHTIGGSVALRASRIAGLFSRGPNQN
1812-8-11_M13uni_--_    STHTIGGSVALRASRLAGLFSRGPNQN
1812-8-4_M13uni_--_1    STHTIGGSVALRASGIAGLFSRGPNQN
           Consensus    STHT!GGSvA..AsgiaslfSrGP.QN
```

6

2

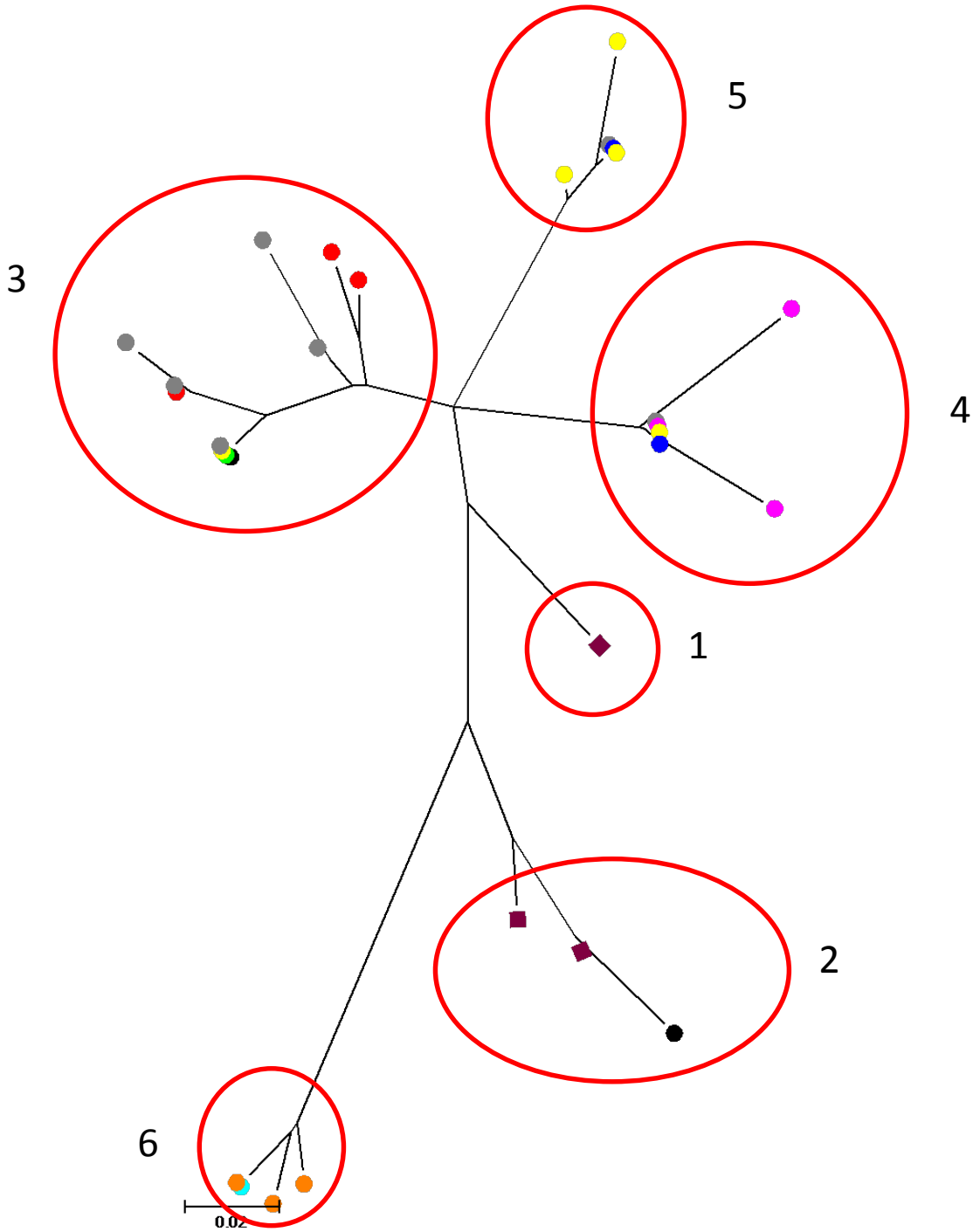
3

1

4

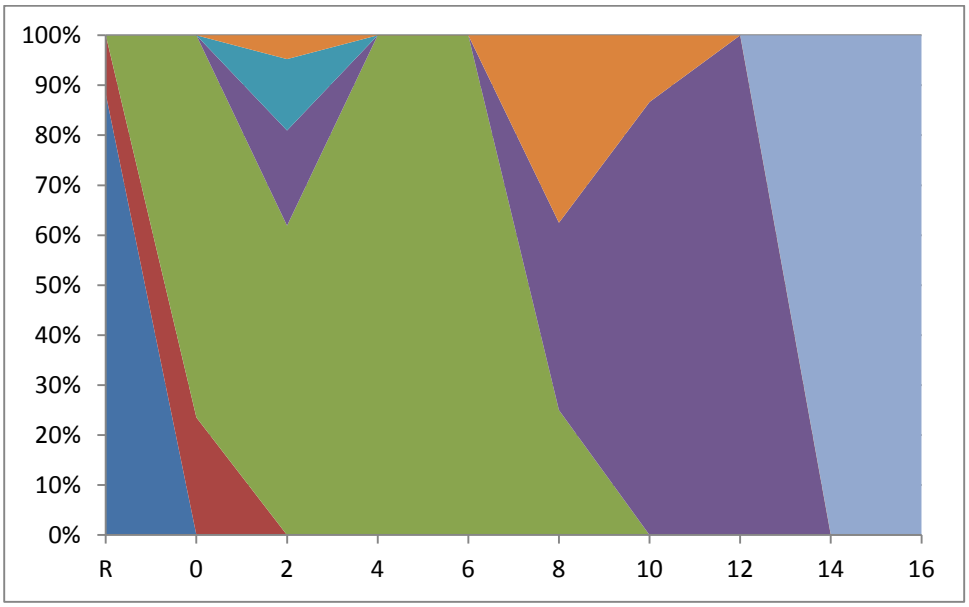
5

O.1Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



O.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

Visualisation of the tree suggests that subpopulations 3,4 and 5 may have arisen through mutation of subpopulation 1.Subpopulation 6 which has replaced all other subpopulations by the completion of the study appears likely to have arisen from subpopulation 2.

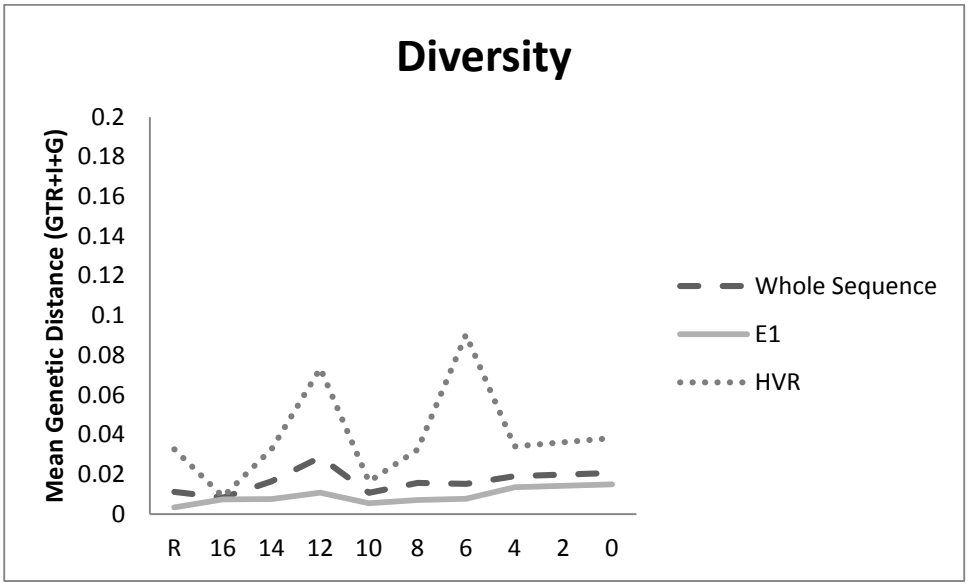


O.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

This provides a clear illustration of sequential change in the subpopulation profile with the elimination of predecessors. This is suggestive of immune mediated HVR1 change.

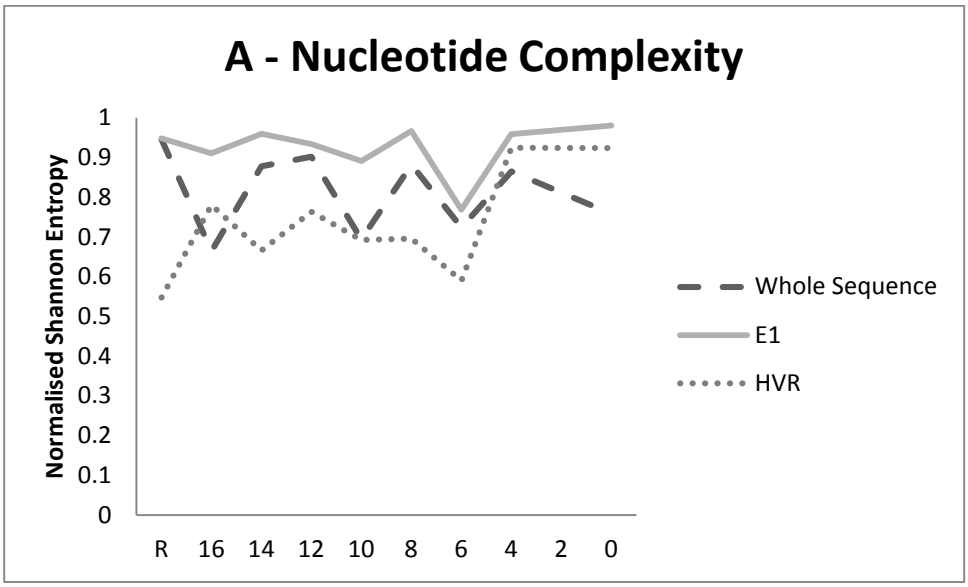
R.1 Subject R

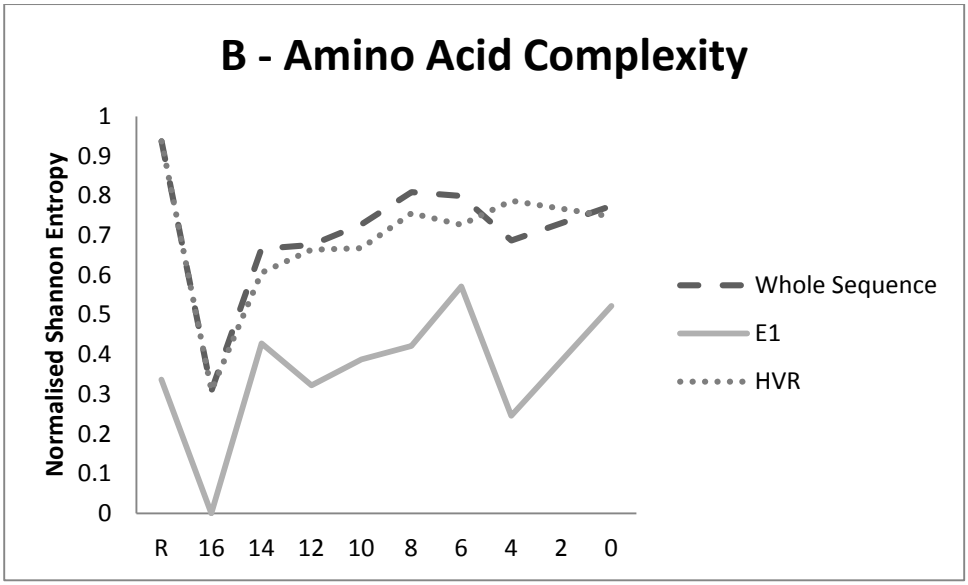
R.1 Diversity, Complexity, and Divergence



R.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

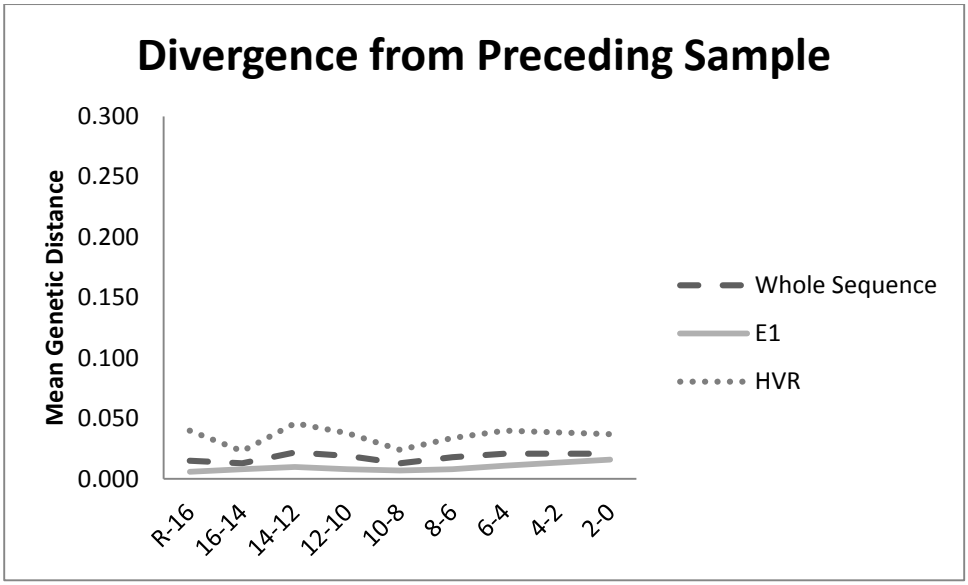
HVR1 diversity is greater than E1 diversity.





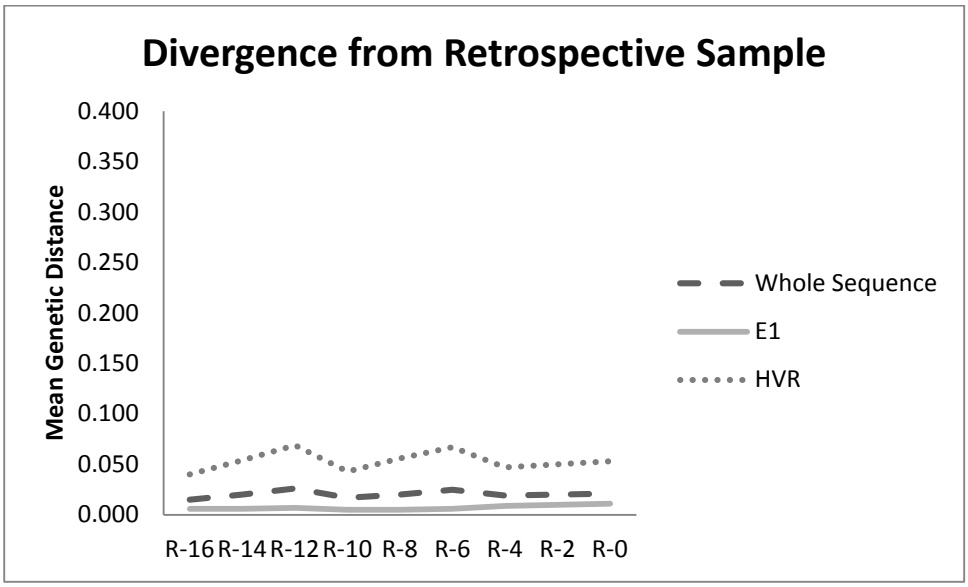
R.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates increased amino acid complexity relative to E1 throughout the study period but less nucleotide complexity.



R.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

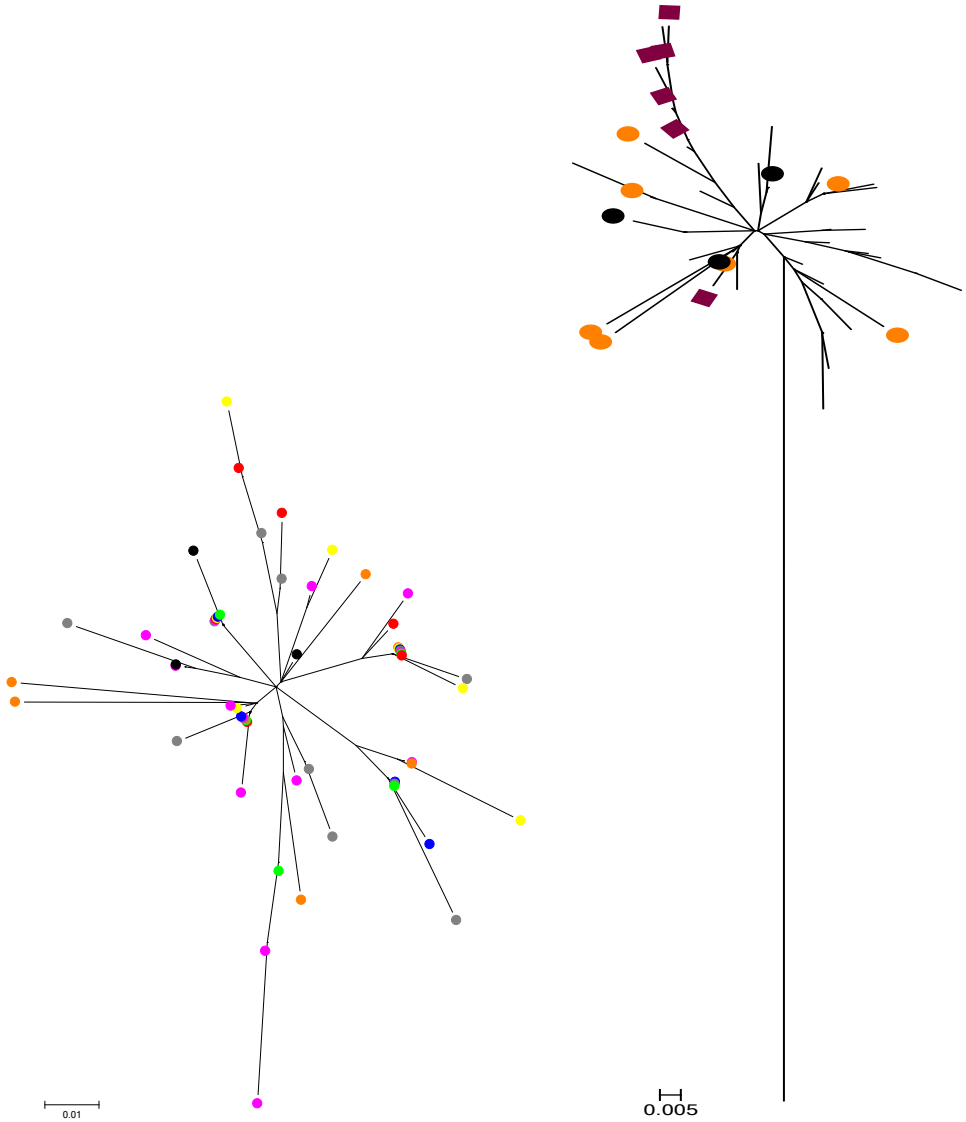
It is notable that despite the longer time interval between the retrospective sample and the intervals between the remaining study samples which corresponds to two weeks that there is a similar magnitude of divergence. Overall, there is little HVR1 or E1 divergence during the study.



R.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

Overall, there is little HVR1 or E1 divergence during the study.

R.1 Phylogenetic analysis



R.1Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (649 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree has been unaffected by the inclusion of the retrospective sample. An outlier red sample from week 12 is remote from the remainder of the tree but analysis of E1 characteristics indicates that it is not a contaminant. The retrospective sequences cluster closely among the remainder of the sequences.

R.1 Subpopulation analysis

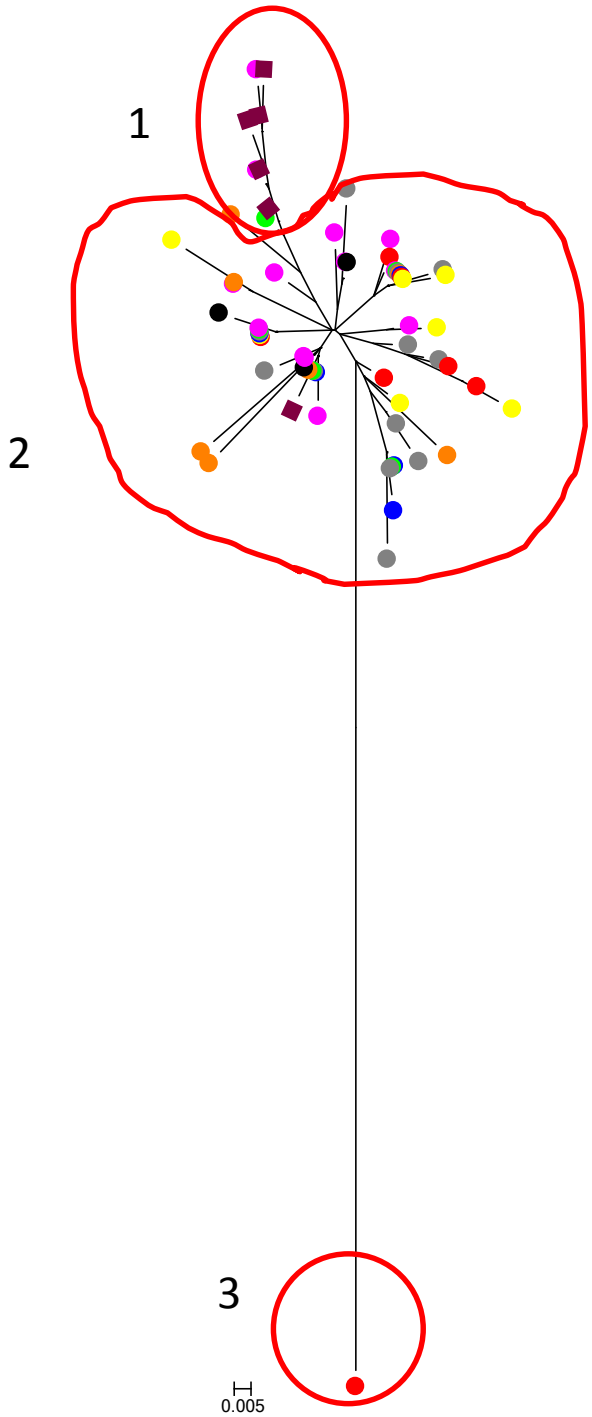
```
                1        10        20      27
                |--------+---------+------|
   0501-0-6     NTYVSGGAAARSAHSLTSLFSLGPNQN
          R     NTYVSGGAPARSAHSLTSLFSLGAQQN
     R-4-22     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-4-10     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-0-2     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-4-11     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-0-17     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-4-4     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-4-7     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-4-9     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1023     NTYVSGGAAARSAHSLTSLFSLGAQQN
 0501-12-26     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-101     NTYVSGGAAARSAHSLTSLFSLGAQQN
      R-8-7     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1210     NTYVSGGAAARSAHSLTSLFSLGAQQN
      R-8-3     NTYVSGGAAARSAHSLTSLFSLGAQQN
     R-8-13     NTYVSGGAAARSAHSLTSLFSLGAQQN
     R-8-11     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1420     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-143     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1418     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1216     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1414     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1411     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-1410     NTYVSGGAAARSAHSLTSLFSLGAQQN
  5165-16-3     NTYVSGGAAARSAHSLTSLFSLGAQQN
  5165-16-1     NTYVSGGAAARSAHSLTSLFSLGAQQN
  0501-6-12     NTYVSGGAAARSAHSLTSLFSLGAQQN
   0501-6-7     NTYVSGGAAARSAHSLTSLFSLGAQQN
       R-11     NTYVSGGAAARSAHSLTILFSLGAQQN
   0501-141     NTYVSGGAAARSAHSFTSLFSLGAQQN
  0501-1419     NTYVSGGAAARSAHSFTRLFSLGAQQN
   0501-0-1     NTYATGGAAARSAHSLTSLFSLGAQQN
   0501-4-1     NTYVSGGAAARSAHSLTSLFSLGSKQN
   0501-0-4     NTYVSGGAAARSAHSLTSLFSLGSKQN
  0501-4-13     NTYVSGGAAARSAHSLTSLFSLGSKQN
  0501-1010     NTYVSGGAAARSAHSLTSLFSLGSKQN
     R-8-19     NTYVSGGAAARSAHSLTSLFSLGSKQN
     R-8-15     NTYVSGGAAARSAHSLTSLFSLGSKQN
   0501-146     NTYVSGGAAARSAHSLTSLFSLGSKQN
  0501-0-15     NTYVSGGAAARSAHSLTSLFSLGSKQN
  0501-1211     NTYVSGGAAARSAHSLTSLFSLGSKQN
  0501-1413     NTYVSGGAAARSAHSLTSLFSLGSKQN
  0501-1218     NTYVSGGAAARSAHSLTSLFSLGSKQN
 5165-16-20     NTYVSGGAAARSAHSLTSLFSLGSKQN
   0501-6-4     NTYVSGGAAARSAHSLTSLFSLGSKQN
   0501-6-1     NTYVSGGAAARSAHSFTSLFSLGSKQN
  0501-1421     NTYVSGGAAARSAHSFTSLFSLGSKQN
      R-8-4     NTYVSGGAAARSAHSFTSLFSLGSKQN
  0501-1024     NTYVSGGAAARSAHSFTSLFSLGSKQN
  0501-6-11     NTHVSGGAAARSAHSFTSLFSLGSKQN
  0501-1417     NTYVIGGAAARSAHSFTSLFSLGSNQN
  0501-4-12     DTYVSGGAVARSAHSLTSLFSPGARQN
       R-20     DTYVSGGAVARSAHSLTSLFSPGARQN
       R-10     DTYVSGGAAARSAHSLTSLFSPGARQN
        R-9     DTYVSGGAVARSAHSLTSLFSPGARHN
  0501-4-18     NTYVSGGAVARSAHSLTSLFSPGARQN
       R-14     NTYVSGGAVARSAHSLTSLFSPGARQN
        R-1     NTYVSGGAAARSAHSLTSLFSPGARQN
  0501-1020     NTYVSGGAVARSAHSLTSLFSPGAQQN
  0501-0-10     NTYVSGGAAARSAHSFTSLFSPGAQQN
   0501-121     TTYVTGGSAARSASGLVSLFSLGPKQN
  Consensus     #TYVsGGAaARSAHSlTSLFSlGa.QN
```

1

2

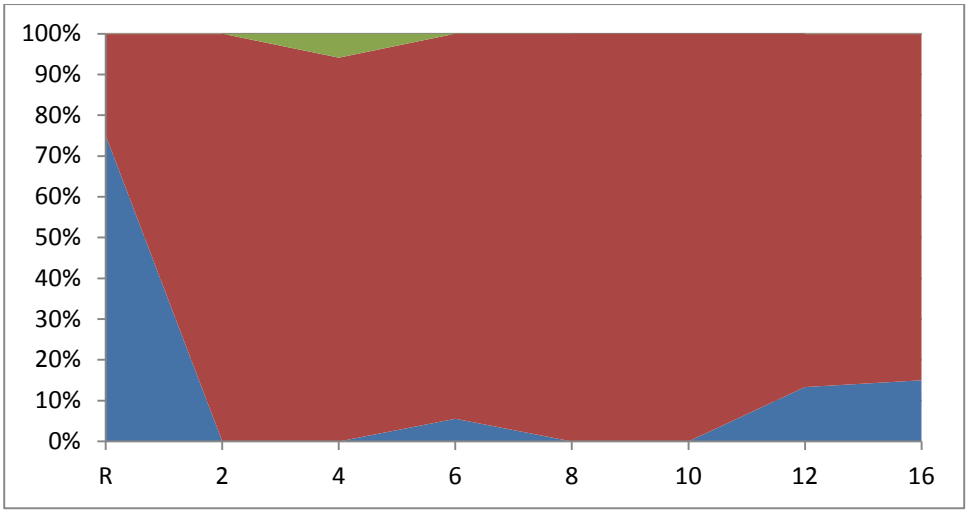
3

R.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



R.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 0 – orange. Identical sequences overlap.

Subpopulation 1 is mostly comprised of retrospective sequences, though sequences within this subpopulation appear intermittently in subsequent samples.

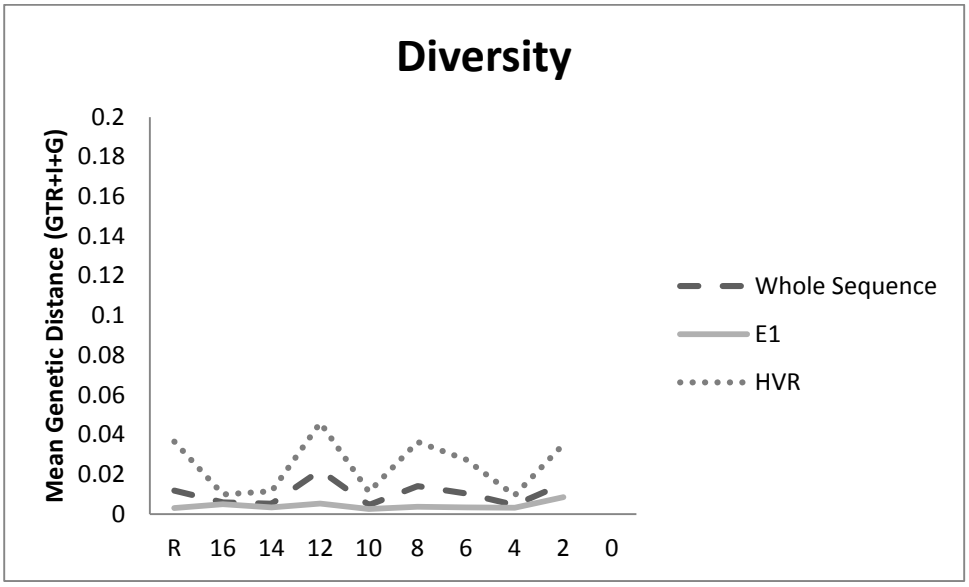


R.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subpopulation 1, initially dominant, is replaced by subpopulation2 which maintains its dominance through the remainder of the study.

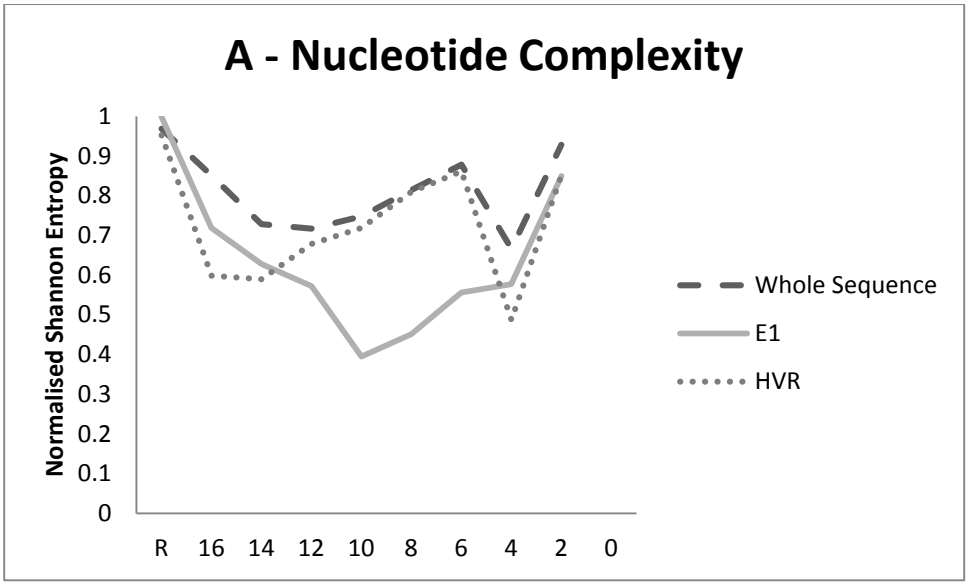
U.1 Subject U

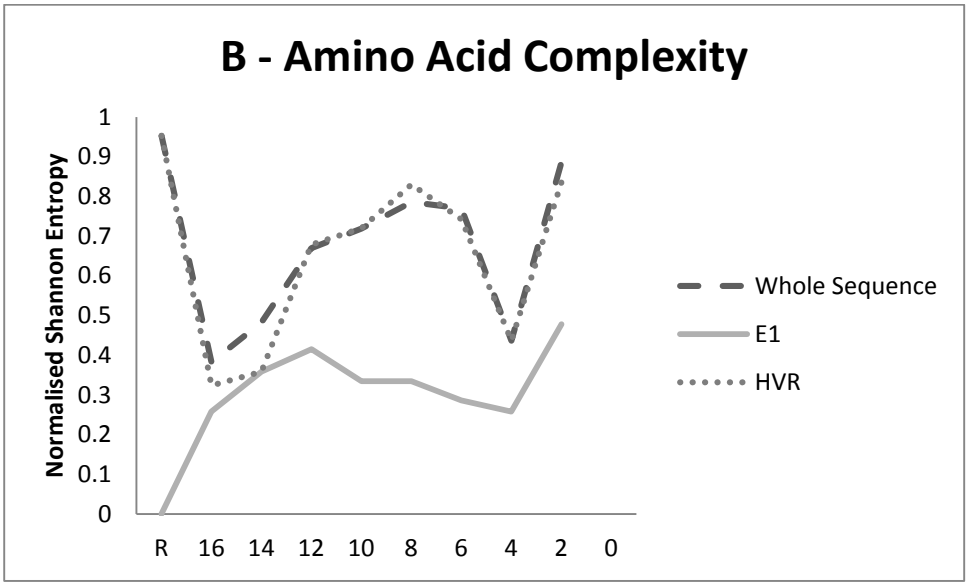
U.1 Diversity, Complexity, and Divergence



U.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

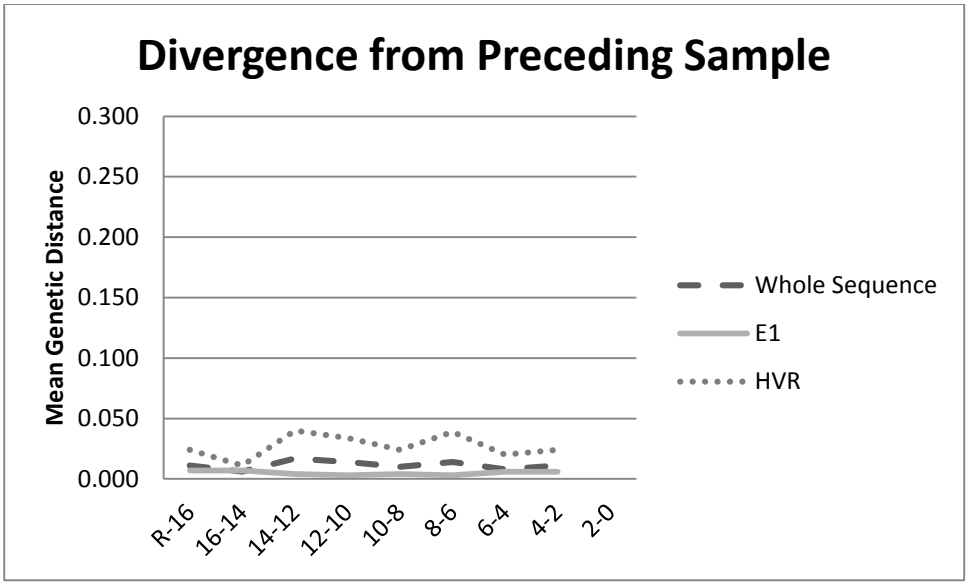
HVR1 diversity is greater than E1 diversity.





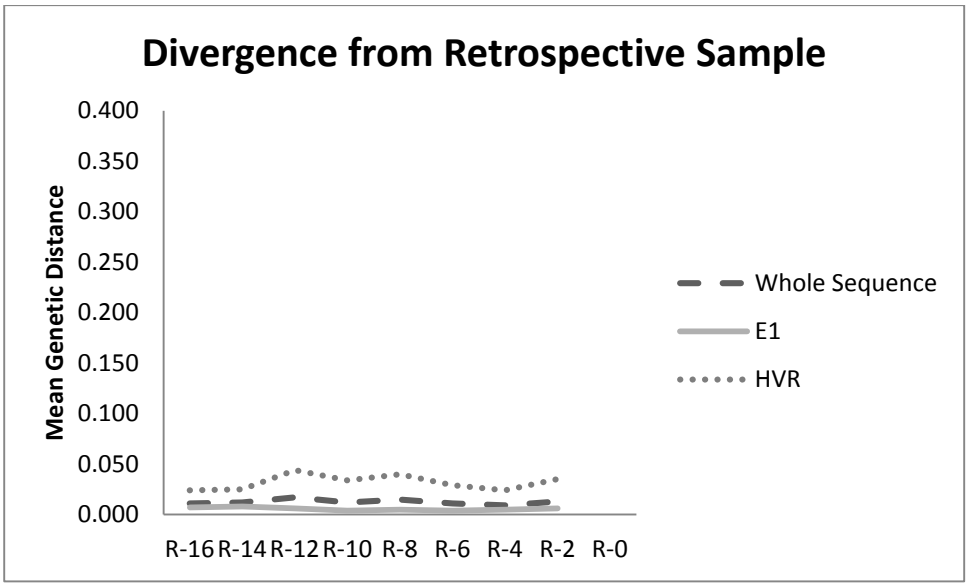
U.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates markedly increased amino acid complexity relative to E1 throughout the study period.



U.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

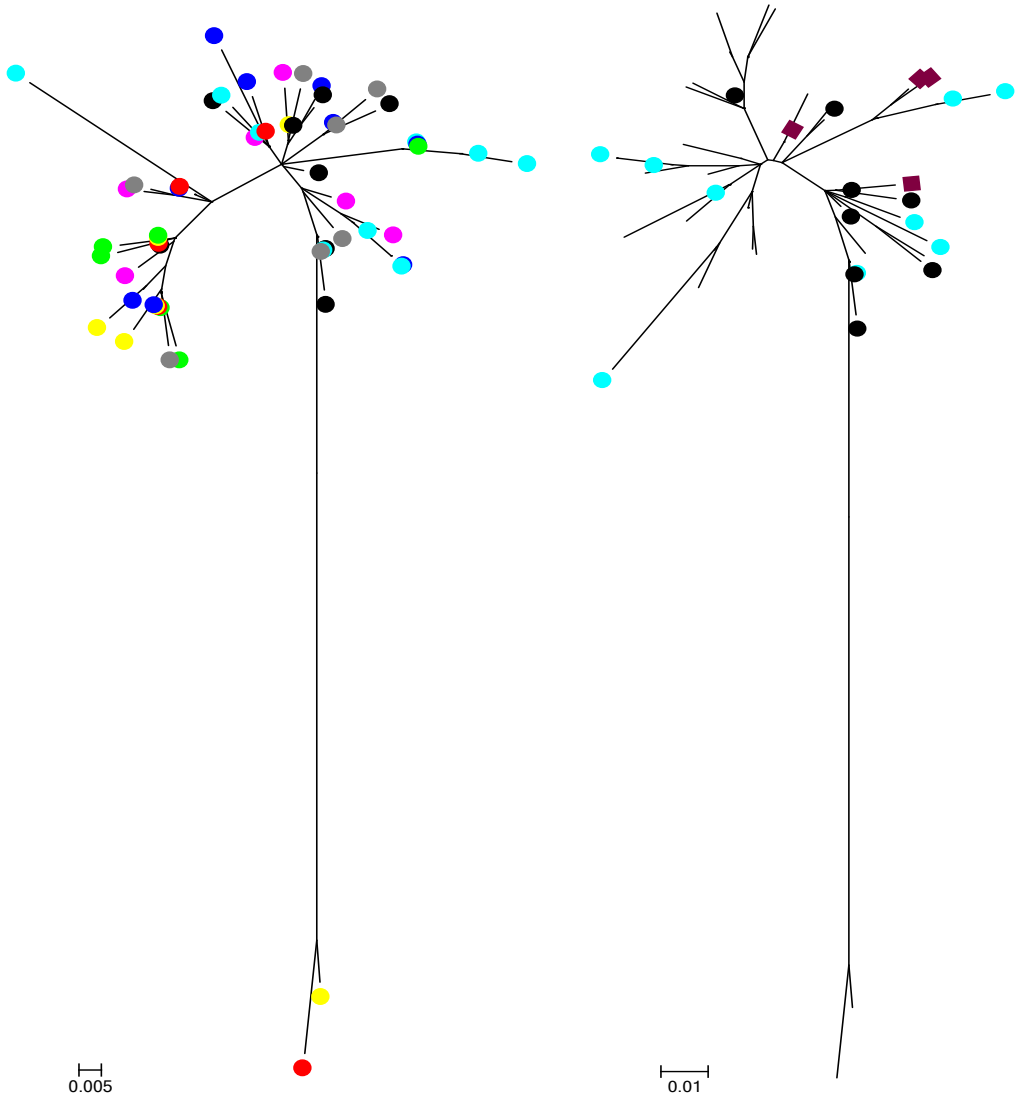
There is little E1 or HVR1 divergence during the study



U.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 and HVR1 demonstrate minimal divergence throughout the study period.

U.1 Phylogenetic analysis



U.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (287 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise. Identical sequences overlap.

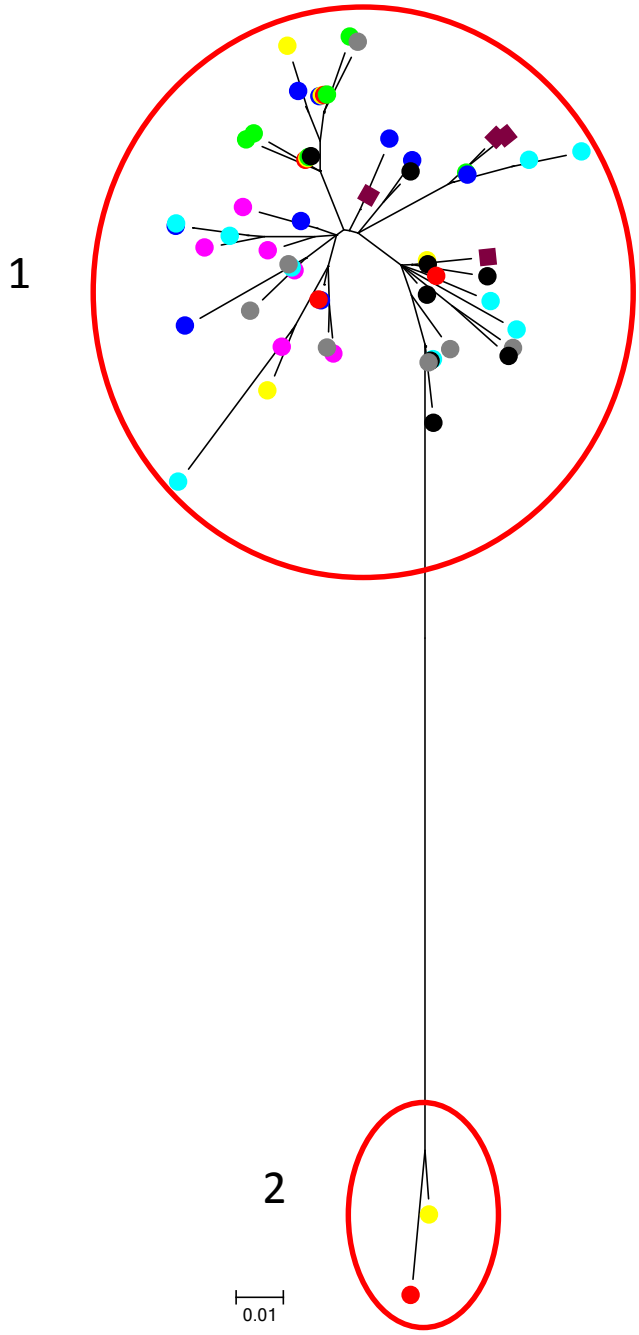
It is noticeable that the general shape of the tree has been unaffected by the inclusion of the retrospective sample. Retrospective sequences cluster among the sequences from the pre treatment samples.

U.1Subpopulation analysis

```
                        1         10         20      27
                        |--------+---------+------|
W-2-1_M13uni_--_11..    TTTISGGTAAYSTRGIASFFAPGARQD
W-4-1_M13uni_--_27..    TTTISGGTAAYSTRGIASFFAPGARQD
W-4-7_M13uni_--_26..    TTTISGGTAAYSTRGIASFFAPGARQD
W-2-19_M13uni_--_13.    TTTISGGTAAYSTRGIASFFAPGARQD
W-6-1_M13uni_--_11..    TTTISGGTAAYSTRGIASFFAPGARQD
W-6-14_M13uni_--_21.    TTTISGGTAAYSTRGIASFFAPGARQD
W-6-19_M13uni_--_9..    TTTISGGTAAYSTRGIASFFAPGARQD
W-13_M13uni-21_--_10    TTTISGGTAAYSTRGIASFFAPGARQD
W-16-2_M13uni_--_19.    TTTISGGTAAYSTRGIASFFAPGARQD
W-2-23_M13uni_--_59.    TTTISGGTAAYSTRGIASFFAPGARQD
W-16-5_M13uni_--_9..    TTTISGGTAAYSTRGIASFFAPGARQD
W-16-20_M13uni_--_11    TTTISGGTAAYSTRGIASFFAPGARQD
W-16-15_M13uni_--_13    TTTISGGTAAYSTRGIASFFAPGARQD
W-16-10a_M13uni_--_1    TTTISGGTAAYSTRGIASFFAPGARQD
W-14-13_M13uni_--_20    TTTISGGTAAYSTRGIASFFAPGARQD
W-14-20_M13uni_--_26    TTTISGGTAAYSTRGIASFFAPGARQD
W-14-19_M13uni_--_9.    TTTISGGTAAYSTRGIASFFAPGARQD
W-14-10_M13uni_--_8.    TTTISGGTAAYSTRGIASFFAPGARQD
W-14-1_M13uni_--_12.    TTTISGGTAAYSTRGIASFFAPGARQD
W-8-19_M13uni_--_16.    TTTISGGTAAYSTRGIASFFAPGARQD
W-12-11_M13uni_--_16    TTTISGGTAAYSTRGIASFFAPGARQD
W-4-17_M13uni_--_18.    TTTISGGTAAYNTRGIASFFAPGARQD
W-6-2_M13uni_--_14..    TTTISGGTAAYNTRGIASFFAPGARQD
W-12-19_M13uni_--_13    TTTISGGTAAYNTRGIASFFAPGARQD
W-16-10_M13uni_--_19    TTTISGGTTAYSTRGIASFFAPGARQD
W-14-16_M13uni_--_10    TTTISGGTAAYSTRGIASFFVPGARQD
W-6-17_M13uni_--_18.    TTTISGGTATYSTRGIASFFAPGARQD
W-22_M13uni-21_--_29    TTTISGGTAAYSTRGIVSFFAPGARQD
W-4-23_M13uni_--_28.    TTTISGGTAAYSTRGIASLFAPGARQD
W-16-16_M13uni_--_8.    TTTISGGTAAYSTRGIASIFAPGARQD
W-6-21_M13uni_--_22.    TTTISGGTAAYSTRGIASFFAPGAMQN
W-2-11_M13uni_--_8..    TTTISGGTAAYSTRGIASFFAPGAMQD
W-4-8_M13uni_--_11..    TTTISGGTAAYNTRGIASIFAPGARQD
W-2-15_M13uni_--_19.    TTTISGGTAAYNTRGIASIFAPGARQD
W-2-21_M13uni_--_315    TTTISGGTAAYNTRGIASIFAPGARQD
W-6-20_M13uni_--_26.    TTTISGGTAAYNTRGIASIFAPGARQD
W-16-11_M13uni_--_12    TTTISGGTAAYNTRGIASIFAPGARQD
W-6-3_M13uni_--_22..    TTTISGGTAAYNTRGIASIFAPGARQD
W-8-1_M13uni_--_13..    TTTISGGTAAYNTRGIASIFAPGARQD
W-12-1_M13uni_--_88.    TTTISGGTAAYNTRGIASIFAPGARQD
W-10-24_M13uni_--_20    TTTISGGTAAYNTRGIASIFAPGARQD
W-10-10_M13uni_--_13    TTTISGGTAAYNTRGIASIFAPGARQD
W-10-1_M13uni_--_19.    TTTISGGTAAYNTRGIASIFAPGARQD
W-2-5_M13uni_--_19..    TTTVSGGTAAYNTRGIASIFAPGARQD
W-6-10_M13uni_--_8..    TTTISGGTAAYNTRGLASIFAPGARQD
W-12_M13uni-21_--_51    TTTISGGTAAYNTRGLASIFAPGARQD
W-8-10_M13uni_--_14.    TTTISGGTAAYNTRGLASIFAPGARQD
W-12-8_M13uni_--_9..    TTTISGGTAAYNTRGLASIFAPGARQD
W-8-20_M13uni_--_16.    TTTISGGTAAYNTRGLASIFAPGARQD
W-8-21_M13uni_--_14.    TTTISGGTAAYNTRGLASIFAPGARQD
W-10-11_M13uni_--_16    TTTISGGTAAYNTRGLASIFAPGARQD
W-21_M13uni-21_--_57    TTTISGGTAAYNTRGFASIFAPGARQD
W-10-6_M13uni_--_14.    TTTISGGTAAYNSRGLASIFAPGARQD
W-2-22_M13uni_--_23.    TTTISGGTAAYNTRGIASIFAPGARQN
W-14-23_M13uni_--_12    TTTTSGGTAAYNTRGLASIFAPGARQD
W-4-18_M13uni_--_22.    TTTISGGTAAYSTRGIASLFAPGAMQD
W-10-19_M13uni_--_24    TTTISGVTAAYNTRGIASIFAPGARQD
W-8-2_M13uni_--_14..    TTTISGGTAARSTYGITSLFSAGPRQN
W-12-14_M13uni_--_9.    TTTISGGTAARSTYGITSLFSAGPRQN
           Consensus    TTTISGGTAAysTrGiaS.FapGarQ#
```

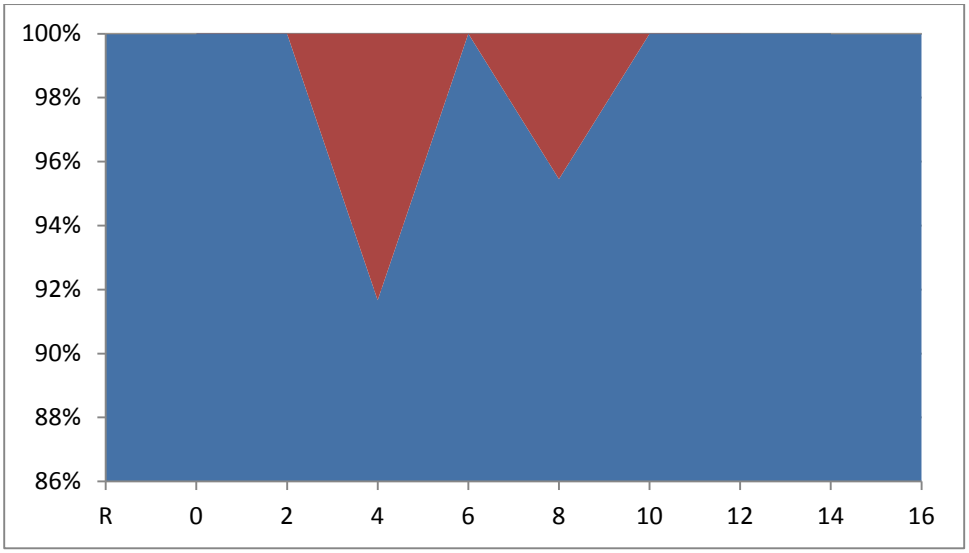
1

U.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



U.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise. Identical sequences overlap.

A majority of sequences are included in subpopulation 1.

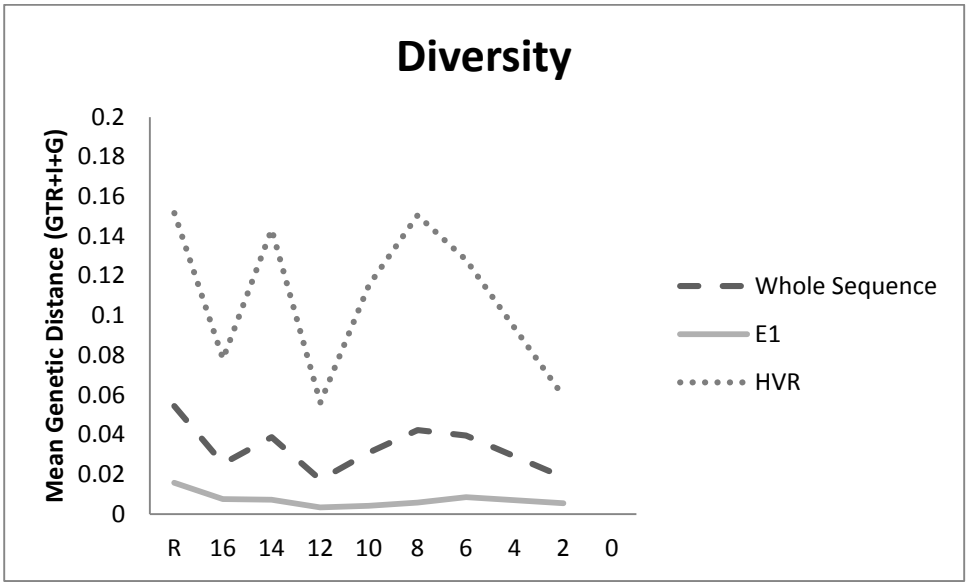


U.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subpopulation 1 comprises >90% of sequences at all times during the study.

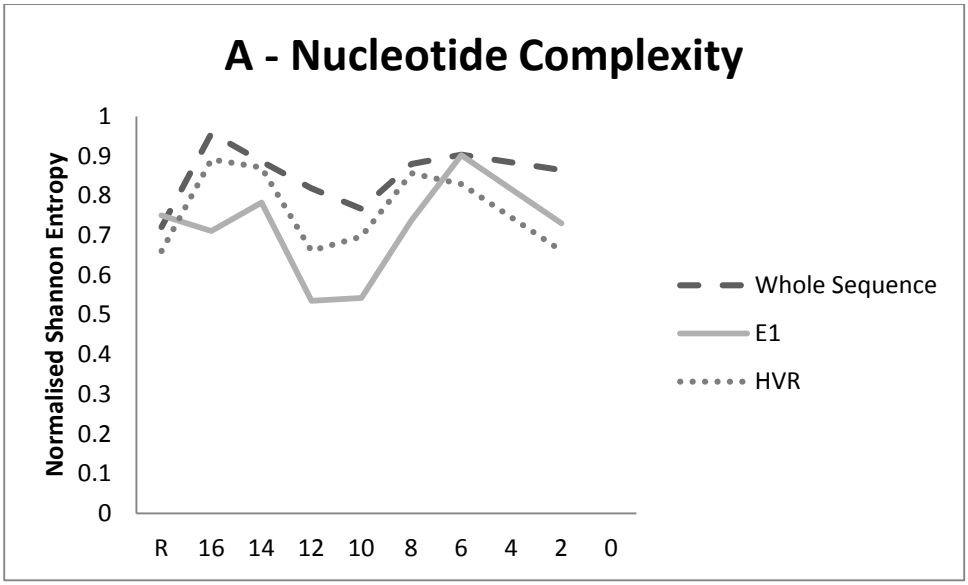
V.1 Subject V

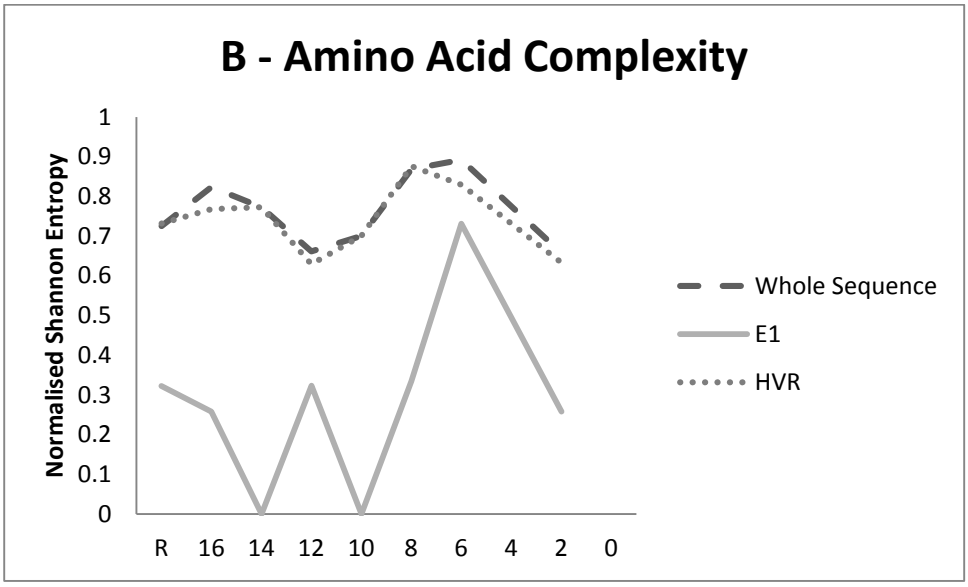
V.1 Diversity, Complexity, and Divergence



V.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

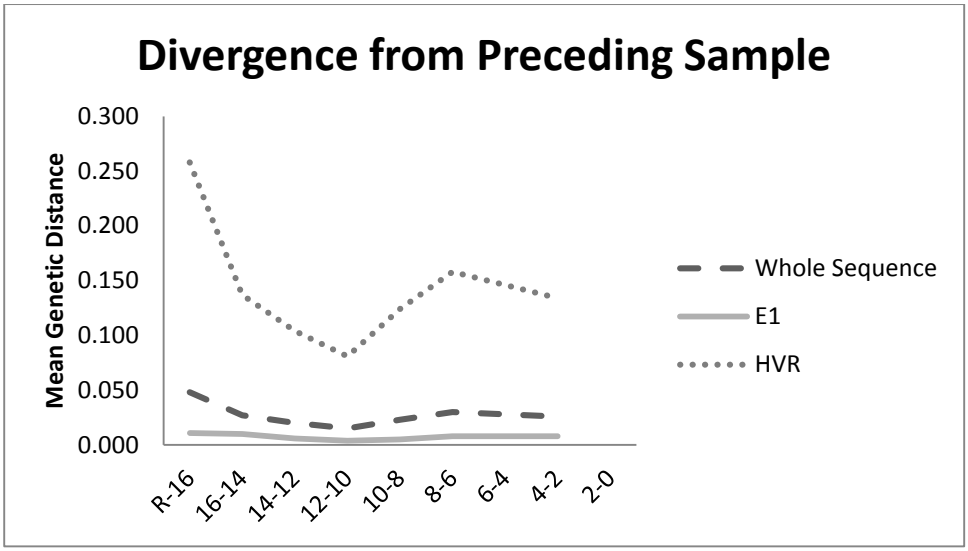
HVR1 diversity is far greater than E1 diversity.





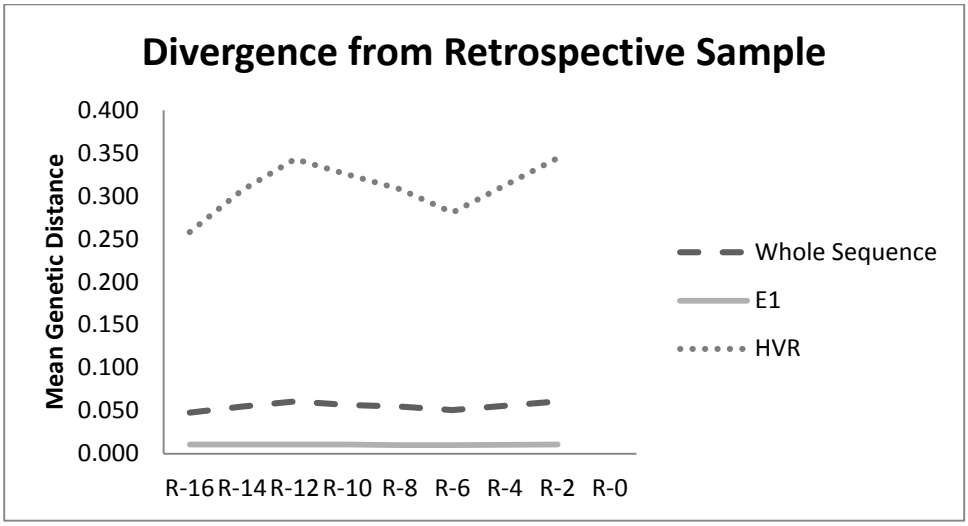
V.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates markedly increased amino acid complexity relative to E1 throughout the study period.



V.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

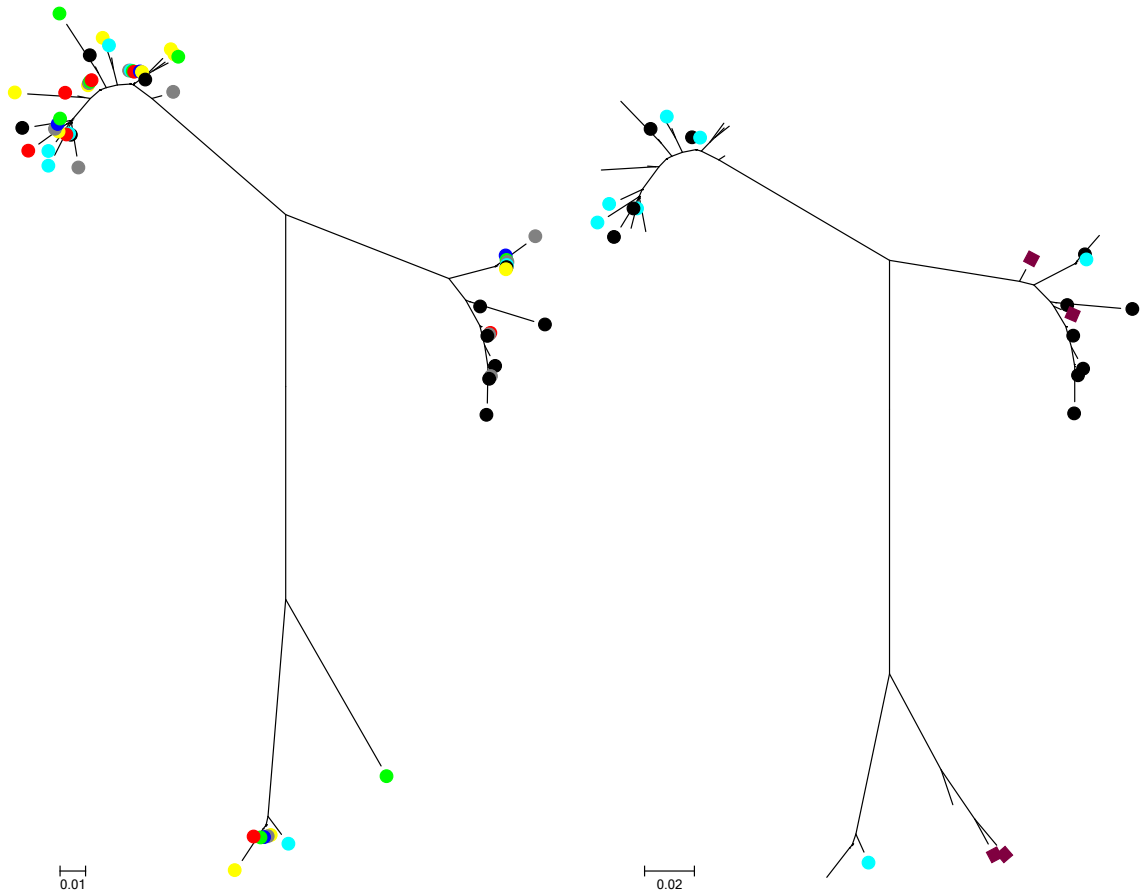
HVR1 divergence is maximal between the retrospective sample and the week 16 samples which is a fifteen month interval. The subsequent fortnightly HVR1 divergence is also significant. E1 demonstrates not significant divergence through the entire study period.



V.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 demonstrates minimal divergence throughout the study period. HVR1 divergence from the retrospective group increases throughout the study period and is maximal at the pre treatment sample

V.1 Phylogenetic analysis



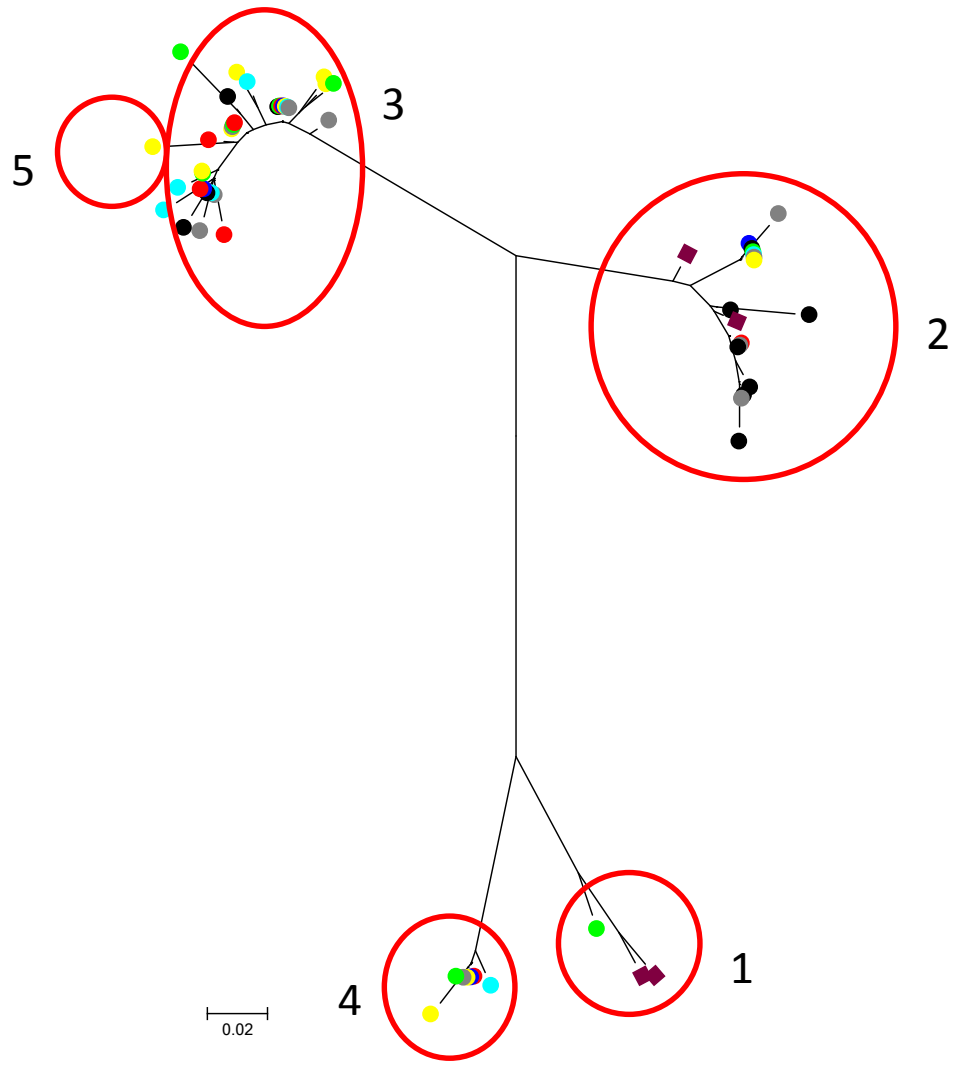
V.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (462 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 2 – turquoise. Identical sequences overlap.

It is noticeable that the general shape of the tree has been unaffected by the inclusion of the retrospective sample. The retrospective sequences cluster in two of the clades that were identified in the samples leading up to treatment.

Subpopulation analysis

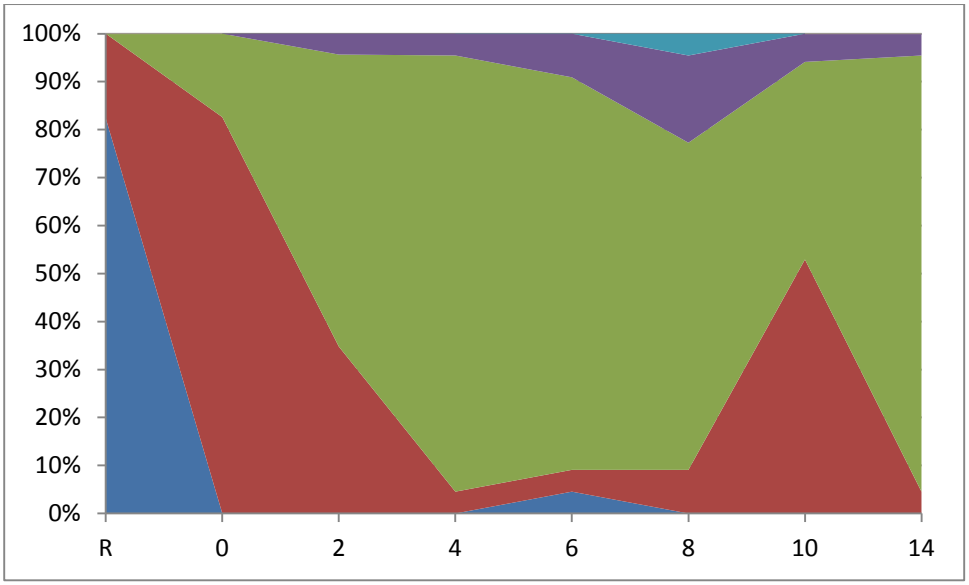
```
                         1         10         20      27
                         |--------+---------+------|
Y-2-1_M13uni_--_12..     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-6-17_M13uni_--_9..     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-8-1_M13uni_--_13..     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-16-12_M13uni_--_7.     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-10-1_M13uni_--_12.     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-14-24_M13uni_--_26     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-12-1_M13uni_--_9..     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-14-1_M13uni_--_17.     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-12-15_M13uni_--_15     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-2-15_M13uni_--_9..     TTYTSGGTVARGASTLAGLFTLGPSQK
Y-16-13_M13uni_--_21     TTFTSGGTVARGASTLAGLFTLGPSQK
Y-8-17_M13uni_--_10.     TTYTSGGSVARGASTLAGLFTLGPSQK
Y-14-22_M13uni_--_22     TTYTSGGSVARGASTLAGLFTLGPSQK
Y-12-2_M13uni_--_10.     TTYTSGGSVARGASTLAGLFTLGPSQK
Y-8-21_M13uni_--_8..     TTYTSGGSVARGASTLAGLFTLGPSQK        3
Y-10-10_M13uni_--_12     TTYTSGGSVARGASTLAGLFTLGPSQK
Y-2-3_M13uni_--_7..9     TTYASGGTVARGASTLAGLFTLGPSQK
Y-6-12_M13uni_--_8..     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-16-22_M13uni_--_10     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-8-10_M13uni_--_22.     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-2-23_M13uni_--_25.     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-8-5_M13uni_--_20..     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-10-19_M13uni_--_20     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-10-23_M13uni_--_8.     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-2-14_M13uni_--_8..     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-14-17_M13uni_--_17     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-12-11_M13uni_--_22     TTYTSGGSVARGASTLTGLFTLGPSQK
Y-16-3_M13uni_--_7..     TTYTSGGAVARGASTLTGLFTLGPSQK
Y-8-24_M13uni_--_20.     TTYTSGGSVARGASTLTGLFTLGPSQN
Y-10-16_M13uni_--_23     TTYTSGGVVARGASTLSGLFTLGPSQK
Y-8-19_M13uni_--_8..     TTYTSGGTVARGASTLAGLFTRGPSQN        5
Y-12-22_M13uni_--_12     TTYTSGGTVARGASTLAGLFTLGPSQN
Y-14-12_M13uni_--_17     TTYTSGGSVARGASTLTRLFTLGPSQK
Y-6-1_M13uni_--_10..     TTYTTGGSAAYGVRSFTSLFTLGPSQN
Y-8-20_M13uni_--_10.     TTYTTGGSAAYGVRSFTSLFTLGPSQN
Y-2-21_M13uni_--_8..     TTYTTGGSAAYGVRSFTSLFTLGPSQN
Y-10-20_M13uni_--_16     TTYTTGGSAAYGVRSFTSLFTLGPSQN
Y-14-2_M13uni_--_14.     TTYTTGGSAAYGVRSFTSLFTLGPSQN
Y-16-17_M13uni_--_11     TTYTTGGSAAYGVRSFTSLFTLGPSQN
Y-14-15_M13uni_--_26     TTYTTGGSAAYGVGSFTSLFTLGPSQN
Y-12-23_M13uni_--_10     TTYTTGGSAAYAARSITSLFTLGPSQN        2
Y-14-10_M13uni_--_14     TTYTTGGSAAYAARSITSLFTLGPSQN
Y-14-23_M13uni_--_25     TTYTTGGSAAYAARSITSLFTLGPSQN
Y-16-1_M13uni_--_10.     TTYTTGGSAAYAARSITSLFTLGPSQN
Y-16-11_M13uni_--_20     TTYTTGGSAAYAARSITSLFTLGPSQN
Y-16-2_M13uni_--_11.     TTYTTGGSAAYAARSLTSLFTLGPSQN
Y-16-24_M13uni_--_9.     TTYTTGGSAAYAARSLTSLFTLGPSQN
Y-18_M13uni-21_--_39     TTYTTGGSAAYAARSFTSLFTLGPSQN
Y-11_M13uni-21_--_83     TTYTTGGSAAYAARSFTSLFTLGPSQN
Y-16-15_M13uni_--_9.     TTYTTGGSAAYAARSITSFFTLGPSQN
Y-16-10_M13uni_--_8.     TTYTTGGSAAYAARSITSLFTLGPSQK
Y-6-14_M13uni_--_10.     ETHVTGGSVAHTVSGLSGLFTRGPSQN
Y-8-12_M13uni_--_18.     ETHVTGGSVAHTVSGLSGLFTRGPSQN
Y-8-15_M13uni_--_12.     ETHVTGGSVAHTVSGLSGLFTRGPSQN        4
Y-10-13_M13uni_--_15     ETHVTGGSVAHTVSGLSGLFTRGPSQN
Y-12-20_M13uni_--_11     ETHVTGGSVAHTVSGLSGLFTRGPSQN
Y-14-19_M13uni_--_16     ETHVTGGSVAHTVSGLSGLFTRGPSQN
Y-2-10_M13uni_--_8..     ETHVTGGSAAHTVSGLSGLFTRGPSQN
Y-10-2_M13uni_--_22.     ETHVTGGSAAHNARRLTSLFSIGPSQN        1
Y-1_M13uni-21_--_25.     ETHVTGGSAAHTARGLTSLFSKGPSQN
Y-10_M13uni-21_--_12     ETRVTGGSAAHTARGLTSLFSQGPSQN
           Consensus     tTyttGGsvA.gas.ltgLFtlGPSQn
```

V.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



V.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 6 – blue, Week 4 – pink, Week 2 – turquoise. Identical sequences overlap.

The retrospective sequences group with two previously identified subpopulations.



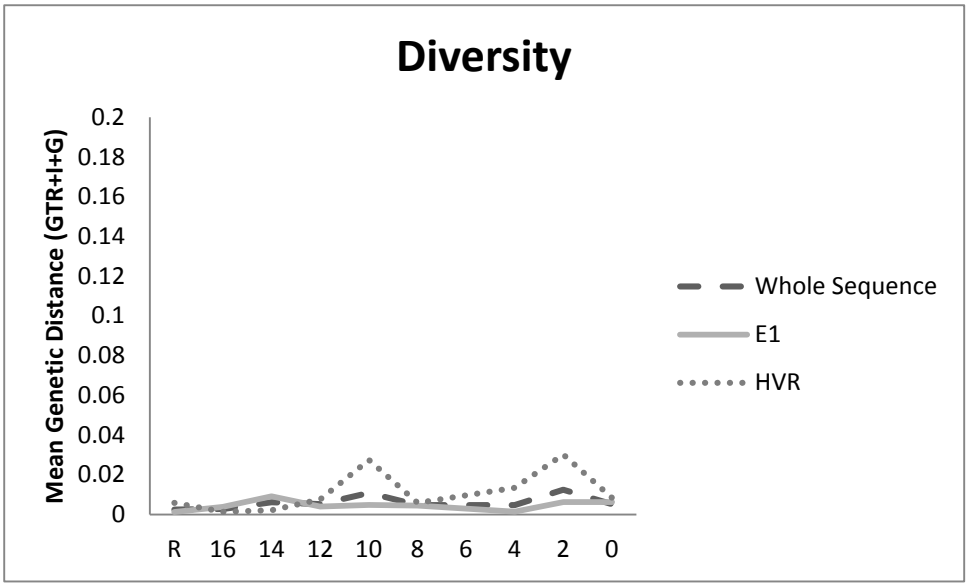
V.1 Fig 8.The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

This demonstrates characteristic features of a time order phylogeny with 3 different dominant HVR1

QS subpopulations identified during the study.

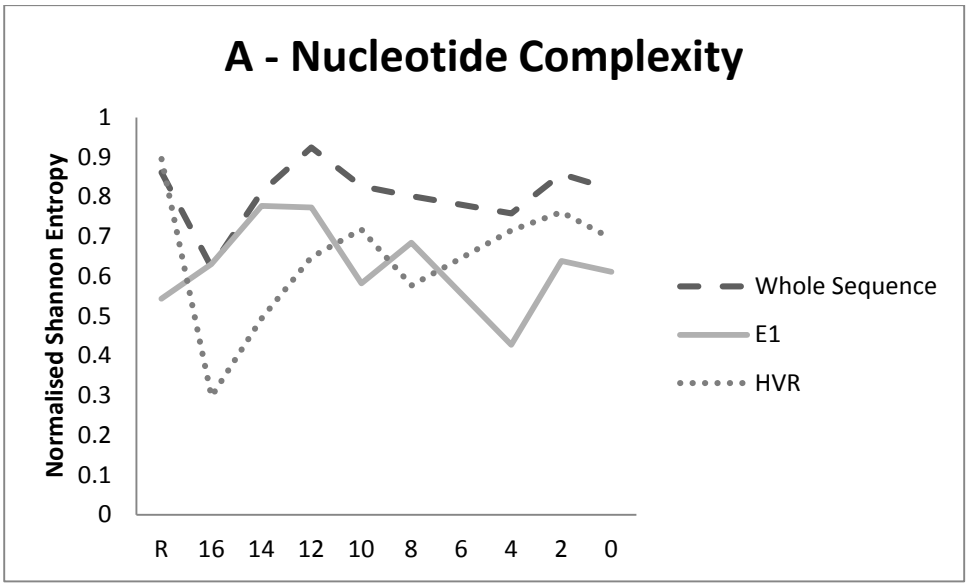
W.1 Subject W

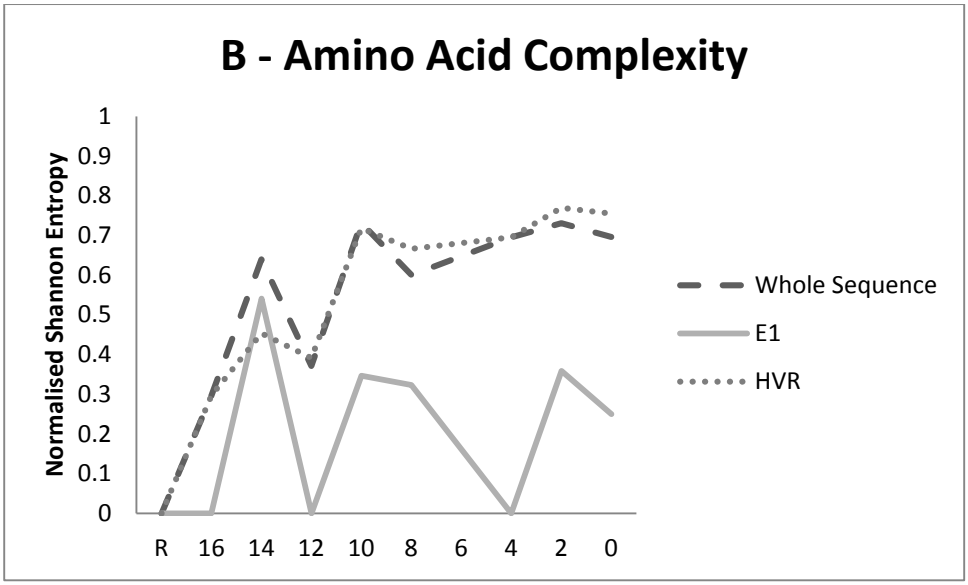
W.1 Diversity, Complexity, and Divergence



W.1 Fig 1. HVR1 QS Diversity for each sample. Diversity is mean pairwise substitutions between clones within the sample and was calculated using a generalised time reversible model with invariant sites and a gamma distribution (GTR+I+G).

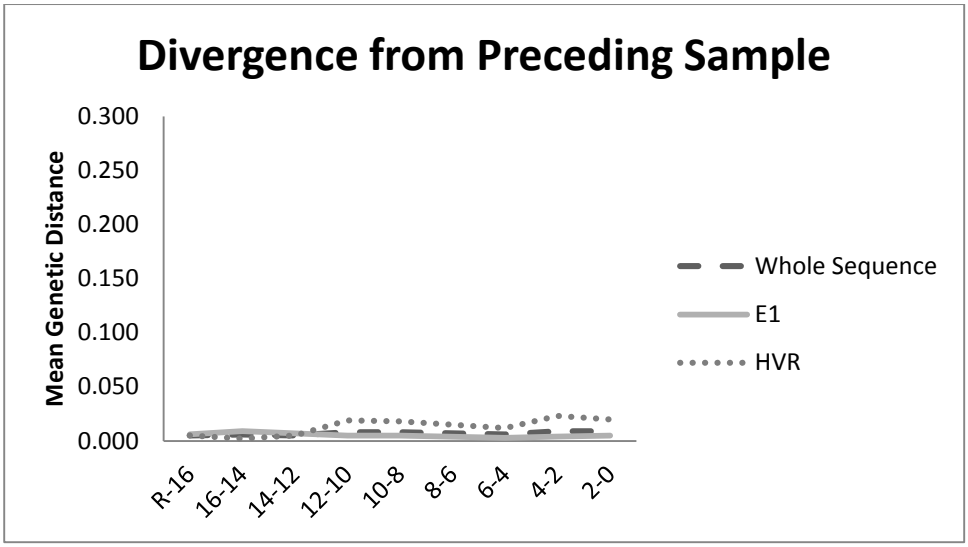
HVR1 diversity is similar to E1 diversity in most samples.





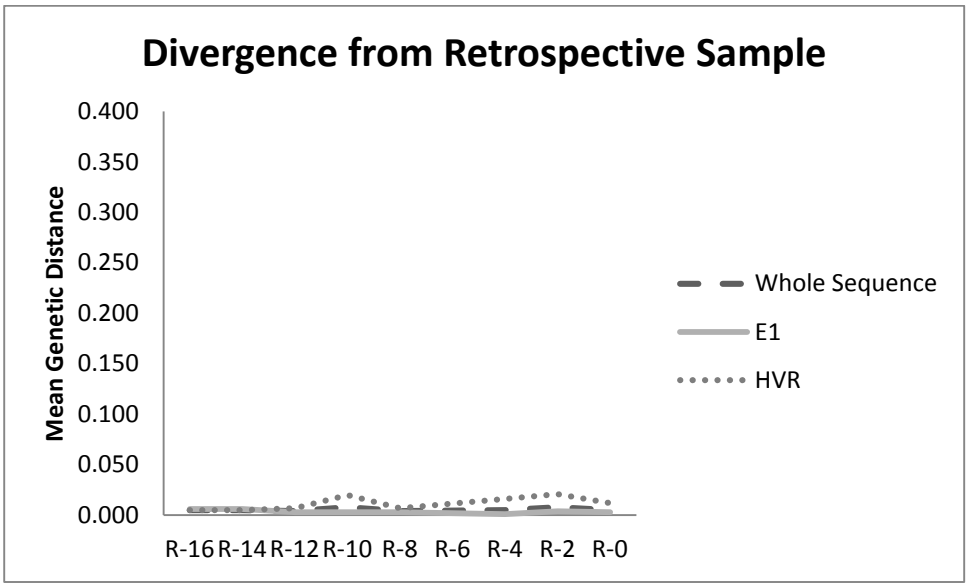
W.1 Fig 2. QS complexity at (A)nucleotide and (B)amino acid level as calculated using Normalised Shannon Entropy.

HVR1 demonstrates markedly increased amino acid complexity relative to E1 throughout the study period.



W.1. Fig 3. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each subsequent group of clones.

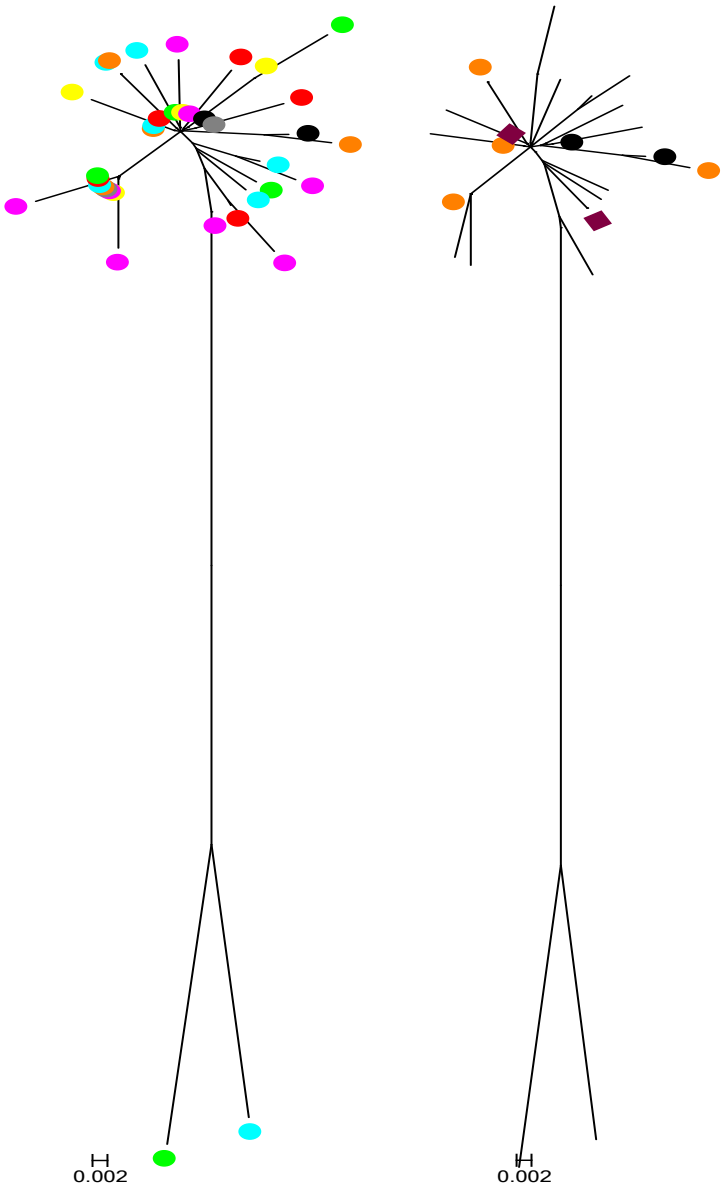
Notably, there is minimal E1 or HVR1 divergence throughout the study which corresponds to fifteen months prior to commencement of treatment.



W.1. Fig 4. QS divergence as measured using gamma distributed maximum composite likelihood pairwise analysis of transitions and transversions between each group of clones and the retrospective groups of clones.

E1 and HVR1 demonstrate minimal divergence throughout the study period.

W.1 Phylogenetic analysis



W.1 Fig. 5. Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment. Right - Phylogenetic tree (left) including all unique HVR1 sequences for the 16 weeks pre treatment with the addition of the unique HVR1 sequences from the retrospective sample (332 days prior to Week 16 sample). Retrospective (wine) and samples from week 16 (black) and week 0 (orange) labelled. Tree constructed using maximum composite likelihood with GTR+I+G and bootstrap 10,000 for the purposes of optimisation. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

It is noticeable that the general shape of the tree has been unaffected by the inclusion of the retrospective sample.

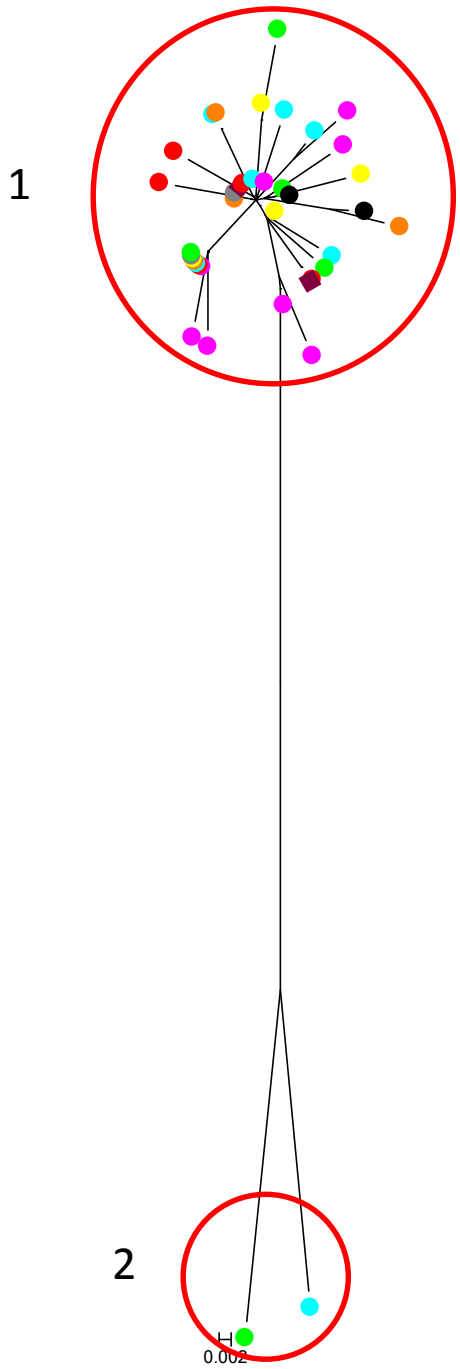
W.1 Subpopulation analysis

```
                        1        10         20      27
                        |--------+---------+------|
Z-0-1_M13uni_--_21..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-2-14_M13uni_--_17.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-0-16_M13uni_--_11.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-2-15_M13uni_--_16.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-2_M13uni-21_--_13.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-10_M13uni-21_--_45    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-2-24_M13uni_--_18.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z16-1_M13uni-21_--_9    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-2-6_M13uni_--_15..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-14-1_M13uni_--_10.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z12-23_M13uni-21_--_    TTYITGGSTARVTQSLVGLFSQGSRQN
Z12-24_M13uni-21_--_    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-2-7_M13uni_--_18..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z12-13_M13uni-21_--_    TTYITGGSTARVTQSLVGLFSQGSRQN
Z12-11_M13uni-21_--_    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-4-18_M13uni_--_9..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-8-1_M13uni_--_25..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-8-3_M13uni_--_59..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-8-5_M13uni_--_45..    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-10-12_M13uni_--_7.    TTYITGGSTARVTQSLVGLFSQGSRQN
Z-10-3_M13uni_--_343    TTYITGGSTARVTQSLVGLFSQGSKQN
Z-2-1_M13uni_--_13..    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-0-11_M13uni_--_18.    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-10-14_M13uni_--_47    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-8-10_M13uni_--_12.    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-14-19_M13uni_--_23    TTYITGGSSARVTQSLVGLFSQGSRQN
Z12-22_M13uni-21_--_    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-4-1_M13uni_--_11..    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-4-21_M13uni_--_9..    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-4-14_M13uni_--_9..    TTYITGGSSARVTQSLVGLFSQGSRQN
Z-4-22_M13uni_--_19.    TTYITGGSTARATQSLVGLFSQGSRQN
Z-10-21_M13uni_--_21    TTYITGGSTAHVTQSLVGLFSQGSRQN
Z-4-24_M13uni_--_23.    TTYTTGGSTARVTQSLVGLFSQGSRQN
Z-4-5_M13uni_--_13..    TTYTTGGSTARVTQSLVGLFSQGSRQN
Z16-15_M13uni-21_--_    TTYITGGSTARVTHSLVGLFSQGSRQN
Z-0-9_M13uni_--_22..    TTYITGGSSARVTHSLVGLFSQGSRQN
Z-4-11_M13uni_--_9..    TTYITGGSSARVTQRLVGLFSQGSRQN
Z-2-5_M13uni_--_11..    ETYTTAGSTARFTQSFVNLFSPGSRQN
Z-10-16_M13uni_--_41    ETYTTAGSTARFTQSFVSLFNPGSKQN
           Consensus    tTYiTgGStARvTQSlVgLFSqGSrQN
```

1

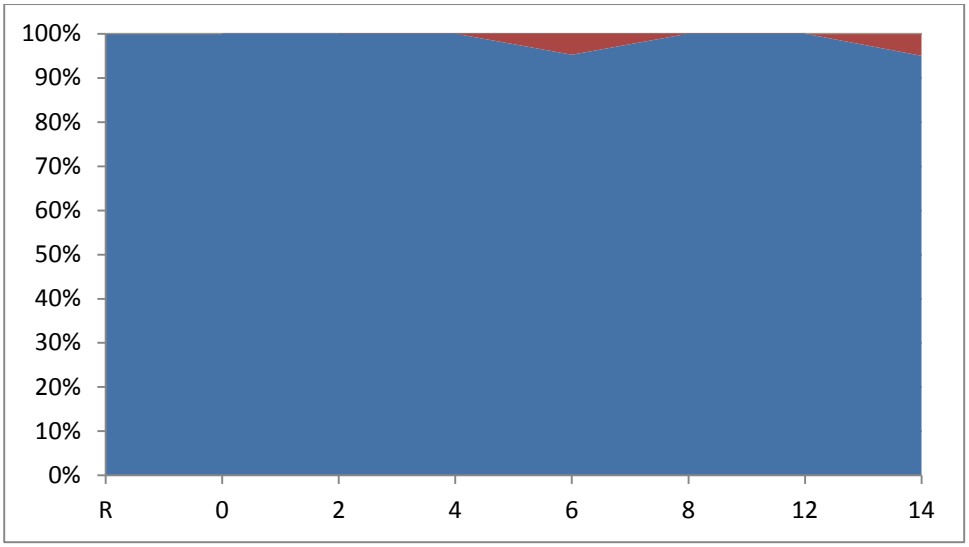
2

W.1 Fig 6 Sequence alignment generated using multalin (http://multalin.toulouse.inra.fr) containing all unique amino acid sequences for each sample. The bottom line approximates a HVR1 consensus sequence for the entire study. This was used to identify HVR1 subpopulations. We defined subpopulations as groups of sequences that differed from all other sequences for the same subject by a minimum of 4 amino acid substitutions. The subpopulations identified (4 in total) are designated by red integers. The numbering of subpopulations was done in accordance with the temporal appearance of the first of each subpopulation. Where two subpopulations appeared in the same sample, the subpopulation which contained the higher number of sequences was labelled first.



W.1 Fig 7 Phylogenetic tree with all unique HVR1 sequences including the retrospective sample with the subpopulations as identified using multalin labelled and circled in red. The labels are: Retrospective clones – wine, Week 16 – black, Week 14 – grey, Week 12 – red, Week 10 – green, Week 8 – yellow, Week 4 – pink, Week 2 – turquoise, Week 0 – orange. Identical sequences overlap.

The inclusion of the retrospective sample does not alter the number of subpopulations.



W.1 Fig 8. The prevalence of each subpopulation from the retrospective sample through the study period to the pre treatment sample.

Subject W is characterised by a dominant HVR1 QS throughout the study with a minor subpopulation appearing twice at weeks 6 and 14.

Bibliography

1. Krugman S, Giles JP, Hammond J. Infectious hepatitis: Evidence for two distinctive clinical, epidemiological, and immunological types of infection. Jama. 1967;200(5):365-73.
2. Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 1989;244(4902):359-62.
3. Tobler LH, Busch MP. History of posttransfusion hepatitis. Clinical chemistry. 1997;43(8):1487-93.
4. Paraná R, Vitvitski L, Andrade Z, Trepo C, Cotrim H, Bertillon P, et al. Acute sporadic non-A, non-B hepatitis in Northeastern Brazil: Etiology and natural history. Hepatology. 1999;30(1):289-93.
5. Flores R, Owens R, Mahy B, Van Regenmortel M. Encyclopedia of Virology. Encyclopedia of Virology. 2008.
6. Rümenapf T, Thiel HJ. Molecular biology of pestiviruses. Animal viruses: molecular biology. 2008:39-96.
7. Thiel H, Collett M, Gould E, Heinz F, Houghton M, Meyers G, et al. Family flaviviridae. Virus taxonomy, VIII report of the International Committee on Taxonomy of Viruses, Academic Press, San Diego. 2005:981-98.
8. Beames B, Chavez D, Lanford RE. GB virus B as a model for hepatitis C virus. ILAR Journal. 2001;42(2):152-60.
9. Chen Z, Benureau Y, Rijnbrand R, Yi J, Wang T, Warter L, et al. GB virus B disrupts RIG-I signaling by NS3/4A-mediated cleavage of the adaptor protein MAVS. Journal of virology. 2007;81(2):964.
10. Martin A, Sangar DV, Lemon SM, Rijnbrand R. Chimeric gb virus b (gbv-b). Google Patents; 2009.
11. Bhattarai N, Stapleton JT. GB virus C: the good boy virus? Trends in microbiology. 2012;20(3):124-30.
12. Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, et al. A proposed system for the nomenclature of hepatitis C viral genotypes. Hepatology. 1994;19(5):1321-4.
13. Simmonds P, Bukh J, Combet C, Deléage G, Enomoto N, Feinstone S, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology. 2005;42(4):962-73.
14. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology. 2014;59(1):318-27.
15. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. The Lancet. 2001;358(9286):958-65.
16. Pang PS, Planet PJ, Glenn JS. The evolution of the major hepatitis C genotypes correlates with clinical response to interferon therapy. PloS one. 2009;4(8):e6579.
17. Patton HM, Patel K, Behling C, Bylund D, Blatt LM, Vallée M, et al. The impact of steatosis on disease progression and early and sustained treatment response in chronic hepatitis C patients. Journal of hepatology. 2004;40(3):484-90.
18. Buckton AJ, Ngui S-L, Arnold C, Boast K, Kovacs J, Klapper PE, et al. Multitypic hepatitis C virus infection identified by real-time nucleotide sequencing of minority genotypes. Journal of clinical microbiology. 2006;44(8):2779-84.
19. Schröter M, Feucht H-H, Zöllner B, Schäfer P, Laufs R. Multiple infections with different HCV genotypes: prevalence and clinical impact. Journal of Clinical Virology. 2003;27(2):200-4.
20. Foster GR, Hézode C, Bronowicki JP, Carosi G, Weiland O, Verlinden L, et al. Telaprevir alone or with peginterferon and ribavirin reduces HCV RNA in patients with chronic genotype 2 but not genotype 3 infections. Gastroenterology. 2011;141(3):881-9. e1.
21. Afdhal N, Zeuzem S, Kwo P, Chojkier M, Gitlin N, Puoti M, et al. Ledipasvir and sofosbuvir for untreated HCV genotype 1 infection. New England Journal of Medicine. 2014;370(20):1889-98.

22. Gane E, Hyland R, An D, Pang P, Symonds W, Mchutchison J, et al. O6 Sofosbuvir/ledipasvir fixed dose combination is safe and effective in difficult-to-treat populations including genotype-3 patients, decompensated genotype-1 patients, and genotype-1 patients with prior sofosbuvir treatment experience. Journal of hepatology. 2014;1(60):S3-S4.
23. Poordad F, Hezode C, Trinh R, Kowdley KV, Zeuzem S, Agarwal K, et al. ABT-450/r–ombitasvir and dasabuvir with ribavirin for hepatitis C with cirrhosis. New England Journal of Medicine. 2014;370(21):1973-82.
24. Liver EAfSo. EASL Clinical Practice Guidelines: management of hepatitis C virus infection. Journal of hepatology. 2014;60(2):392.
25. Markov PV, Pepin J, Frost E, Deslandes S, Labbe A-C, Pybus OG. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. Journal of General Virology. 2009;90(9):2086-96.
26. Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, Rasachak B, et al. Genetic history of hepatitis C virus in East Asia. Journal of virology. 2009;83(2):1071-82.
27. Kalinina O, Norder H, Mukomolov S, Magnius LO. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. Journal of virology. 2002;76(8):4034-43.
28. Moreau I, Hegarty S, Levis J, Sheehy P, Crosbie O, Kenny-Walsh E, et al. Serendipitous identification of natural intergenotypic recombinants of hepatitis C in Ireland. Virol J. 2006;3:95.
29. Colina R, Casane D, Vasquez S, García-Aguirre L, Chunga A, Romero H, et al. Evidence of intratypic recombination in natural populations of hepatitis C virus. Journal of general virology. 2004;85(1):31.
30. Kageyama S, Agdamag DM, Alesna ET, Leaño PS, Heredia AML, Abellanosa-Tac-An IP, et al. A natural inter-genotypic (2b/1b) recombinant of hepatitis C virus in the Philippines. Journal of medical virology. 2006;78(11):1423-8.
31. Kurbanov F, Tanaka Y, Avazova D, Khan A, Sugauchi F, Kan N, et al. Detection of hepatitis C virus natural recombinant RF1_2k/1b strain among intravenous drug users in Uzbekistan. Hepatology Research. 2008;38(5):457-64.
32. Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. The Lancet infectious diseases. 2005;5(9):558-67.
33. McLeod A, Weir A, Aitken C, Gunson R, Templeton K, Molyneaux P, et al. Rise in testing and diagnosis associated with Scotland's Action Plan on Hepatitis C and introduction of dried blood spot testing. Journal of epidemiology and community health. 2014;68(12):1182-8.
34. Wiessing L, Ferri M, Grady B, Kantzanou M, Sperle I, Cullen KJ, et al. Hepatitis C virus infection epidemiology among people who inject drugs in Europe: A systematic review of data for scaling up treatment and prevention. 2014.
35. Prevost TC, Presanis AM, Taylor A, Goldberg DJ, Hutchinson SJ, De Angelis D. Estimating the number of people with hepatitis C virus who have ever injected drugs and have yet to be diagnosed: an evidence synthesis approach for Scotland. Addiction. 2015.
36. Rein DB, Wittenborn JS, Smith BD, Liffmann DK, Ward JW. The Cost-effectiveness, health benefits, and financial costs of new antiviral treatments for hepatitis C virus. Clinical infectious diseases. 2015:civ220.
37. Razavi H, ElKhoury AC, Elbasha E, Estes C, Pasini K, Poynard T, et al. Chronic hepatitis C virus (HCV) disease burden and cost in the United States. Hepatology. 2013;57(6):2164-70.
38. Pybus OG, Markov PV, Wu A, Tatem AJ. Investigating the endemic transmission of the hepatitis C virus. International journal for parasitology. 2007;37(8):839-49.
39. Power J, Lawlor E, Davidson F, Yap P, Kenny-Walsh E, Whelton M, et al. Hepatitis C viraemia in recipients of Irish intravenous anti-D immunoglobulin. Lancet. 1994;344(8930):1166-7.
40. Power J, Lawlor E, Davidson F, Holmes E, Yap P, Simmonds P. Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin. The Lancet. 1995;345(8959):1211-3.

41. Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. New England Journal of Medicine. 1999;340(16):1228-33.
42. Smith DB, Lawlor E, Power J, O'Riordan J, McAllister J, Lycett C, et al. A Second Outbreak of Hepatitis C Virus Infection from Anti–D Immunoglobulin in Ireland. Vox sanguinis. 1999;76(3):175-80.
43. Wiese M, Grungreiff K, Guthoff W, Lafrenz M, Oesen U, Porst H. Outcome in a hepatitis C (genotype 1b) single source outbreak in Germany--a 25-year multicenter study. Journal of hepatology. 2005;43(4):590-8.
44. Abdel-Aziz F, Habib M, Mohamed MK, Abdel-Hamid M, Gamil F, Madkour S, et al. Hepatitis C virus (HCV) infection in a community in the Nile Delta: population description and HCV prevalence. Hepatology. 2000;32(1):111-5.
45. Frank C, Mohamed MK, Strickland GT, Lavanchy D, Arthur RR, Magder LS, et al. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. The Lancet. 2000;355(9207):887-91.
46. Lehman E, Wilson M. Epidemic hepatitis C virus infection in Egypt: estimates of past incidence and future morbidity and mortality. Journal of viral hepatitis. 2009;16(9):650-8.
47. Vogel M, Deterding K, Wiegand J, Grüner NH, Baumgarten A, Jung MC, et al. Initial presentation of acute hepatitis C virus (HCV) infection among HIV-negative and HIV-positive individuals—experience from 2 large German networks on the study of acute HCV infection. Clinical infectious diseases. 2009;49(2):317-9.
48. Loomba R, Rivera M, McBurney R, Park Y, Haynes-Williams V, Rehermann B, et al. The natural history of acute hepatitis C: clinical presentation, laboratory findings and treatment outcomes. Alimentary Pharmacology & Therapeutics. 2011;33(5):559-65.
49. Mosley JW, Operskalski EA, Tobler LH, Andrews WW, Phelps B, Dockter J, et al. Viral and host factors in early hepatitis C virus infection. Hepatology. 2005;42(1):86-92.
50. Bertoletti A, Ferrari C. Kinetics of the immune response during HBV and HCV infection. Hepatology. 2003;38(1):4-13.
51. Cox AL, Netski DM, Mosbruger T, Sherman SG, Strathdee S, Ompad D, et al. Prospective evaluation of community-acquired acute-phase hepatitis C virus infection. Clinical infectious diseases. 2005;40(7):951-8.
52. Grebely J, Prins M, Hellard M, Cox AL, Osburn WO, Lauer G, et al. Hepatitis C virus clearance, reinfection, and persistence, with insights from studies of injecting drug users: towards a vaccine. The Lancet infectious diseases. 2012;12(5):408-14.
53. Hayashi J, Kishihara Y, Ueno K, Yamaji K, Kawakami Y, Furusyo N, et al. Age-related response to interferon alfa treatment in women vs men with chronic hepatitis C virus infection. Archives of internal medicine. 1998;158(2):177-81.
54. Micallef J, Kaldor J, Dore G. Spontaneous viral clearance following acute hepatitis C infection: a systematic review of longitudinal studies. Journal of viral hepatitis. 2006;13(1):34-41.
55. Rao HY, Sun DG, Jiang D, Yang RF, Guo F, Wang JH, et al. IL28B genetic variants and gender are associated with spontaneous clearance of hepatitis C virus infection. Journal of viral hepatitis. 2012;19(3):173-81.
56. Grebely J, Raffa JD, Lai C, Krajden M, Conway B, Tyndall MW. Factors associated with spontaneous clearance of hepatitis C virus among illicit drug users. Canadian Journal of Gastroenterology. 2007;21(7):447.
57. Syed GH, Amako Y, Siddiqui A. Hepatitis C virus hijacks host lipid metabolism. Trends in Endocrinology & Metabolism. 2010;21(1):33-40.
58. Hui JM, Sud A, Farrell GC, Bandara P, Byth K, Kench JG, et al. Insulin resistance is associated with chronic hepatitis C and virus infection fibrosis progression. Gastroenterology. 2003;125(6):1695-704.
59. Zhu N, Khoshnan A, Schneider R, Matsumoto M, Dennert G, Ware C, et al. Hepatitis C virus core protein binds to the cytoplasmic domain of tumor necrosis factor (TNF) receptor 1 and enhances TNF-induced apoptosis. Journal of virology. 1998;72(5):3691-7.

60. Kato N, Ikeda M, Sugiyama K, Mizutani T, Tanaka T, Shimotohno K. Hepatitis C virus population dynamics in human lymphocytes and hepatocytes infected in vitro. Journal of General Virology. 1998;79(8):1859-69.
61. Forton DM, Karayiannis P, Mahmud N, Taylor-Robinson SD, Thomas HC. 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. Journal of virology. 2004;78(10):5170-83.
62. El–Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology. 2007;132(7):2557-76.
63. Planas R, Ballesté B, Alvarez MA, Rivera M, Montoliu S, Galeras JA, et al. Natural history of decompensated hepatitis C virus-related cirrhosis. A study of 200 patients. Journal of hepatology. 2004;40(5):823-30.
64. Veldt BJ, Heathcote EJ, Wedemeyer H, Reichen J, Hofmann WP, Zeuzem S, et al. Sustained virologic response and clinical outcomes in patients with chronic hepatitis C and advanced fibrosis. Annals of internal medicine. 2007;147(10):677-84.
65. Sangiovanni A, Prati GM, Fasani P, Ronchi G, Romeo R, Manini M, et al. The natural history of compensated cirrhosis due to hepatitis C virus: A 17-year cohort study of 214 patients. Hepatology. 2006;43(6):1303-10.
66. Omland LH, Krarup H, Jepsen P, Georgsen J, Harritshøj LH, Riisom K, et al. Mortality in patients with chronic and cleared hepatitis C viral infection: a nationwide cohort study. Journal of hepatology. 2010;53(1):36-42.
67. Forman LM, Lewis JD, Berlin JA, Feldman HI, Lucey MR. The association between hepatitis C infection and survival after orthotopic liver transplantation. Gastroenterology. 2002;122(4):889-96.
68. Bukh J, Purcell RH, Miller RH. Sequence analysis of the 5'noncoding region of hepatitis C virus. Proceedings of the National Academy of Sciences. 1992;89(11):4942.
69. Friebe P, Lohmann V, Krieger N, Bartenschlager R. Sequences in the 5'nontranslated region of hepatitis C virus required for RNA replication. Journal of virology. 2001;75(24):12047.
70. Otto GA, Puglisi JD. The pathway of HCV IRES-mediated translation initiation. Cell. 2004;119(3):369-80.
71. Luo G, Xin S, Cai Z. Role of the 5'-proximal stem-loop structure of the 5'untranslated region in replication and translation of hepatitis C virus RNA. Journal of virology. 2003;77(5):3312.
72. Sun HY, Ou NY, Wang SW, Liu WC, Cheng TF, Shr SJ, et al. Novel Nucleotide and Amino Acid Covariation between the 5′ UTR and the NS2/NS3 Proteins of Hepatitis C Virus: Bioinformatic and Functional Analyses. PloS one. 2011;6(9):e25530.
73. Yamada N, Tanihara K, Takada A, Yorihuzi T, Tsutsumi M, Shimomura H, et al. Genetic organization and diversity of the 3'noncoding region of the hepatitis C virus genome. Virology. 1996;223(1):255-61.
74. Kolykhalov AA, Feinstone SM, Rice CM. Identification of a highly conserved sequence element at the 3'terminus of hepatitis C virus genome RNA. Journal of virology. 1996;70(6):3363-71.
75. Blight KJ, Rice CM. Secondary structure determination of the conserved 98-base sequence at the 3'terminus of hepatitis C virus genome RNA. Journal of virology. 1997;71(10):7345.
76. Yi MK, Lemon SM. 3'nontranslated RNA signals required for replication of hepatitis C virus RNA. Journal of virology. 2003;77(6):3557.
77. Yanagi M, St Claire M, Emerson SU, Purcell RH, Bukh J. In vivo analysis of the 3′ untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. Proceedings of the National Academy of Sciences. 1999;96(5):2291.
78. Friebe P, Boudet J, Simorre JP, Bartenschlager R. Kissing-loop interaction in the 3′ end of the hepatitis C virus genome essential for RNA replication. Journal of virology. 2005;79(1):380-92.
79. Oh JW, Sheu GT, Lai M. Template requirement and initiation site selection by hepatitis C virus polymerase on a minimal viral RNA template. Journal of Biological Chemistry. 2000;275(23):17710.

80. Friebe P, Bartenschlager R. Genetic analysis of sequences in the 3'nontranslated region of hepatitis C virus that are important for RNA replication. Journal of virology. 2002;76(11):5326.
81. Song Y, Friebe P, Tzima E, Junemann C, Bartenschlager R, Niepmann M. The hepatitis C virus RNA 3'-untranslated region strongly enhances translation directed by the internal ribosome entry site. Journal of virology. 2006;80(23):11579.
82. McLauchlan J, Lemberg MK, Hope G, Martoglio B. Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. The EMBO journal. 2002;21(15):3980-8.
83. Harris C, Herker E, Farese RV, Ott M. The hepatitis C virus core protein decreases lipid droplet turnover: A mechanism for core-induced steatosis. Journal of Biological Chemistry. 2011.
84. Majeau N, Fromentin R, Savard C, Duval M, Tremblay MJ, Leclerc D. Palmitoylation of hepatitis C virus core protein is important for virion production. Journal of Biological Chemistry. 2009;284(49):33915.
85. Moriya K, Yotsuyanagi H, Shintani Y, Fujie H, Ishibashi K, Matsuura Y, et al. Hepatitis C virus core protein induces hepatic steatosis in transgenic mice. Journal of general virology. 1997;78(7):1527.
86. PERLEMUTER G, SABILE A, LETTERON P, VONA G, TOPILCO A, CHRÉTIEN Y, et al. Hepatitis C virus core protein inhibits microsomal triglyceride transfer protein activity and very low density lipoprotein secretion: a model of viral-related steatosis. The FASEB journal. 2002;16(2):185-94.
87. Boulant S, Targett-Adams P, McLauchlan J. Disrupting the association of hepatitis C virus core protein with lipid droplets correlates with a loss in production of infectious virus. Journal of general virology. 2007;88(8):2204.
88. Shavinskaya A, Boulant S, Penin F, McLauchlan J, Bartenschlager R. The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. Journal of Biological Chemistry. 2007;282(51):37158-69.
89. Jackel-Cram C, Babiuk LA, Liu Q. Up-regulation of fatty acid synthase promoter by hepatitis C virus core protein: genotype-3a core has a stronger effect than genotype-1b core. Journal of hepatology. 2007;46(6):999-1008.
90. Mohd-Ismail NK, Deng L, Sukumaran SK, Yu VC, Hotta H, Tan YJ. The hepatitis C virus core protein contains a BH3 domain that regulates apoptosis through specific interaction with human Mcl-1. Journal of virology. 2009;83(19):9993.
91. Ruggieri A, Harada T, Matsuura Y, Miyamura T. Sensitization to Fas-mediated apoptosis by hepatitis C virus core protein. Virology. 1997;229(1):68-76.
92. Berg CP, Schlosser SF, Neukirchen DKH, Papadakis C, Gregor M, Wesselborg S, et al. Hepatitis C virus core protein induces apoptosis-like caspase independent cell death. Virology Journal. 2009;6(1):213.
93. Chang ML, Chen JC, Chang MY, Yeh CT, Lin WP, Liang CK, et al. Acute expression of hepatitis C core protein in adult mouse liver: mitochondrial stress and apoptosis. Scandinavian journal of gastroenterology. 2008;43(6):747-55.
94. Moriya K, Fujie H, Shintani Y, Yotsuyanagi H, Tsutsumi T, Ishibashi K, et al. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. Nature medicine. 1998;4(9):1065-7.
95. Wang T, Campbell R, Yi M, Lemon S, Weinman S. Role of Hepatitis C virus core protein in viral-induced mitochondrial dysfunction. Journal of viral hepatitis. 2010;17(11):784-93.
96. Banerjee S, Saito K, Ait-Goughoulte M, Meyer K, Ray RB, Ray R. Hepatitis C virus core protein upregulates serine phosphorylation of insulin receptor substrate-1 and impairs the downstream akt/protein kinase B signaling pathway for insulin resistance. Journal of virology. 2008;82(6):2606.
97. Pazienza V, Clément S, Pugnale P, Conzelman S, Foti M, Mangia A, et al. The hepatitis C virus core protein of genotypes 3a and 1b downregulates insulin receptor substrate 1 through genotype-specific mechanisms. Hepatology. 2007;45(5):1164-71.
98. Doehle BP, Gale Jr M. Innate Immune Evasion Strategies of HCV and HIV. 2012.
99. Tu Z, Pierce RH, Kurtis J, Kuroki Y, Crispe IN, Orloff MS. Hepatitis C virus core protein subverts the antiviral activities of human Kupffer cells. Gastroenterology. 2010;138(1):305-14.

100. Xu Z, Choi J, Yen TSB, Lu W, Strohecker A, Govindarajan S, et al. Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. The EMBO journal. 2001;20(14):3840-8.
101. Eng FJ, Walewski JL, Klepper AL, Fishman SL, Desai SM, McMullan LK, et al. Internal initiation stimulates production of p8 minicore, a member of a newly discovered family of hepatitis C virus core protein isoforms. Journal of virology. 2009;83(7):3104.
102. Cocquerel L, de Beeck AO, Lambot M, Roussel J, Delgrange D, Pillez A, et al. Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. The EMBO journal. 2002;21(12):2893-902.
103. Vieyres G, Thomas X, Descamps V, Duverlie G, Patel AH, Dubuisson J. Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. Journal of virology. 2010;84(19):10159.
104. Goffard A, Callens N, Bartosch B, Wychowski C, Cosset FL, Montpellier C, et al. Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins. Journal of virology. 2005;79(13):8400.
105. Helle F, Vieyres G, Elkrief L, Popescu CI, Wychowski C, Descamps V, et al. Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. Journal of virology. 2010;84(22):11905-15.
106. Kato N, Ootsuyama Y, Ohkoshi S, Nakazawa T, Sekiya H, Hijikata M, et al. Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. Biochemical and biophysical research communications. 1992;189(1):119-27.
107. Op De Beeck A, Voisset C, Bartosch B, Ciczora Y, Cocquerel L, Keck Z, et al. Characterization of functional hepatitis C virus envelope glycoproteins. Journal of virology. 2004;78(6):2994.
108. Krey T, d'Alayer J, Kikuti CM, Saulnier A, Damier-Piolle L, Petitpas I, et al. The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. PLoS pathogens. 2010;6(2):e1000762.
109. Meunier JC, Russell RS, Goossens V, Priem S, Walter H, Depla E, et al. Isolation and characterization of broadly neutralizing human monoclonal antibodies to the e1 glycoprotein of hepatitis C virus. Journal of virology. 2008;82(2):966-73.
110. Keck ZY, Sung VMH, Perkins S, Rowe J, Paul S, Liang TJ, et al. Human monoclonal antibody to hepatitis C virus E1 glycoprotein that blocks virus attachment and viral infectivity. Journal of virology. 2004;78(13):7257.
111. Mazumdar B, Banerjee A, Meyer K, Ray R. Hepatitis C virus E1 envelope glycoprotein interacts with apolipoproteins in facilitating entry into hepatocytes. Hepatology. 2011;54(4):1149-56.
112. El Omari K, Iourin O, Kadlec J, Sutton G, Harlos K, Grimes JM, et al. Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. Nature communications. 2014;5.
113. Falson P, Bartosch B, Alsaleh K, Tews BA, Loquet A, Ciczora Y, et al. Hepatitis C virus envelope glycoprotein E1 forms trimers at the surface of the virion. Journal of virology. 2015;89(20):10333-46.
114. Broering TJ, Garrity KA, Boatright NK, Sloan SE, Sandor F, Thomas WD, et al. Identification and characterization of broadly neutralizing human monoclonal antibodies directed against the E2 envelope glycoprotein of hepatitis C virus. Journal of virology. 2009;83(23):12473-82.
115. Fournillier-Jacob A, Lunel F, Cahour A, Cresta P, Frangeul L, Perrin M, et al. Antibody responses to hepatitis C envelope proteins in patients with acute or chronic hepatitis C. Journal of medical virology. 1996;50(2):159-67.
116. Law M, Maruyama T, Lewis J, Giang E, Tarr AW, Stamataki Z, et al. Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. Nature medicine. 2008;14(1):25-7.
117. Potter JA, Owsianka AM, Jeffery N, Matthews DJ, Keck Z-Y, Lau P, et al. Toward a hepatitis C virus vaccine: the structural basis of hepatitis C virus neutralization by AP33, a broadly neutralizing antibody. Journal of virology. 2012;86(23):12923-32.
118. de Jong YP, Dorner M, Mommersteeg MC, Xiao JW, Balazs AB, Robbins JB, et al. Broadly neutralizing antibodies abrogate established hepatitis C virus infection. Science translational medicine. 2014;6(254):254ra129-254ra129.

119. Helle F, Duverlie G, Dubuisson J. The hepatitis C virus glycan shield and evasion of the humoral immune response. Viruses. 2011;3(10):1909-32.
120. Bankwitz D, Steinmann E, Bitzegeio J, Ciesek S, Friesland M, Herrmann E, et al. Hepatitis C virus hypervariable region 1 modulates receptor interactions, conceals the CD81 binding site, and protects conserved neutralizing epitopes. Journal of virology. 2010;84(11):5751-63.
121. Penin F, Combet C, Germanidis G, Frainais P-O, Deléage G, Pawlotsky J-M. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. Journal of virology. 2001;75(12):5703-10.
122. Tseng CTK, Klimpel GR. Binding of the hepatitis C virus envelope protein E2 to CD81 inhibits natural killer cell functions. The Journal of experimental medicine. 2002;195(1):43.
123. Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Ohkoshi S, Shimotohno K. Hypervariable regions in the putative glycoprotein of hepatitis C virus1. Biochemical and biophysical research communications. 1991;175(1):220-8.
124. Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, Crawford K, et al. 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. 1991;180(2):842-8.
125. Callens N, Ciczora Y, Bartosch B, Vu-Dac N, Cosset F-L, Pawlotsky J-M, et al. Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein e2 contribute to virus entry. Journal of virology. 2005;79(24):15331-41.
126. Saito S, Kato N, Hijikata M, Gunji T, Itabashi M, Kondo M, et al. Comparison of hypervariable regions (HVR1 and HVR2) in positive-and negative-stranded hepatitis C virus RNA in cancerous and non-cancerous liver tissue, peripheral blood mononuclear cells and serum from a patient with hepatocellular carcinoma. International journal of cancer. 1996;67(2):199-203.
127. Guan M, Wang W, Liu X, Tong Y, Liu Y, Ren H, et al. Three different functional microdomains in the hepatitis C virus hypervariable region 1 (HVR1) mediate entry and immune evasion. Journal of Biological Chemistry. 2012;287(42):35631-45.
128. McCaffrey K, Boo I, Poumbourios P, Drummer HE. Expression and characterization of a minimal hepatitis C virus glycoprotein E2 core domain that retains CD81 binding. Journal of virology. 2007;81(17):9584-90.
129. Hofmann WP, Sarrazin C, Kronenberger B, Schönberger B, Bruch K, Zeuzem S. Mutations within the CD81-Binding Sites and Hypervariable Region 2 of the Envelope 2 Protein: Correlation with Treatment Response in Hepatitis C Virus–Infected Patients. Journal of Infectious Diseases. 2003;187(6):982.
130. Yagnik AT, Lahm A, Meola A, Roccasecca RM, Ercole BB, Nicosia A, et al. A model for the hepatitis C virus envelope glycoprotein E2. Proteins: Structure, Function, and Bioinformatics. 2000;40(3):355-66.
131. Cuevas Torrijos JM, Torres Puente M, Jimenez Hernandez N, Bracho Lapiedra MA, Garcia Robles IR, Wrobel B, et al. Genetic variability of hepatitis C virus before and after combined therapy of interferon plus ribavirin. 2008.
132. Troesch M, Meunier I, Lapierre P, Lapointe N, Alvarez F, Boucher M, et al. Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein. Virology. 2006;352(2):357-67.
133. Torres-Puente M, Cuevas JM, Jiménez-Hernández N, Bracho MA, García-Robles I, Wrobel B, et al. Using evolutionary tools to refine the new hypervariable region 3 within the envelope 2 protein of hepatitis C virus. Infection, Genetics and Evolution. 2008;8(1):74-82.
134. Humphreys I, Fleming V, Fabris P, Parker J, Schulenberg B, Brown A, et al. Full-Length Characterization of Hepatitis C Virus Subtype 3a Reveals Novel Hypervariable Regions under Positive Selection during Acute Infection. Journal of virology. 2009;83(22):11456.
135. Carrere-Kremer S, Montpellier-Pala C, Cocquerel L, Wychowski C, Penin F, Dubuisson J. Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. Journal of virology. 2002;76(8):3720.

136. Sakai A, Claire MS, Faulk K, Govindarajan S, Emerson SU, Purcell RH, et al. The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. Proceedings of the National Academy of Sciences. 2003;100(20):11646.
137. Jones CT, Murray CL, Eastman DK, Tassello J, Rice CM. Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. Journal of virology. 2007;81(16):8374.
138. Harada T, Tautz N, Thiel HJ. E2-p7 region of the bovine viral diarrhea virus polyprotein: processing and functional studies. Journal of virology. 2000;74(20):9498.
139. Luik P, Chew C, Aittoniemi J, Chang J, Wentworth P, Dwek RA, et al. The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. Proceedings of the National Academy of Sciences. 2009;106(31):12712.
140. Pavlović D, Neville DCA, Argaud O, Blumberg B, Dwek RA, Fischer WB, et al. The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. Proceedings of the National Academy of Sciences. 2003;100(10):6104.
141. Lin C, Lindenbach BD, Pragai BM, McCourt DW, Rice CM. Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. Journal of virology. 1994;68(8):5063.
142. KATO N. Genome of human hepatitis C virus (HCV): gene organization, sequence diversity, and variation. Microbial & Comparative Genomics. 2000;5(3):129-51.
143. Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. Journal of virology. 1993;67(7):3835.
144. Santolini E, Pacini L, Fipaldini C, Migliaccio G, Monica N. The NS2 protein of hepatitis C virus is a transmembrane polypeptide. Journal of virology. 1995;69(12):7461.
145. Yamaga AK, Ou J. Membrane topology of the hepatitis C virus NS2 protein. Journal of Biological Chemistry. 2002;277(36):33228.
146. Oliver Koch J, Bartenschlager R. Modulation of hepatitis C virus NS5A hyperphosphorylation by nonstructural proteins NS3, NS4A, and NS4B. Journal of virology. 1999;73(9):7138.
147. Neddermann P, Clementi A, De Francesco R. Hyperphosphorylation of the hepatitis C virus NS5A protein requires an active NS3 protease, NS4A, NS4B, and NS5A encoded on the same polyprotein. Journal of virology. 1999;73(12):9984.
148. Tellinghuisen TL, Foss KL, Treadaway J. Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. PLoS pathogens. 2008;4(3):e1000032.
149. Franck N, Le Seyec J, Guguen-Guillouzo C, Erdtmann L. Hepatitis C virus NS2 protein is phosphorylated by the protein kinase CK2 and targeted for degradation to the proteasome. Journal of virology. 2005;79(5):2700-8.
150. Kaukinen P, Sillanpaa M, Kotenko S, Lin R, Hiscott J, Melén K, et al. Hepatitis C virus NS2 and NS3/4A proteins are potent inhibitors of host cell cytokine/chemokine gene expression. Virol J. 2006;3(66):66.
151. Erdtmann L, Franck N, Lerat H, Le Seyec J, Gilot D, Cannie I, et al. The hepatitis C virus NS2 protein is an inhibitor of CIDE-B-induced apoptosis. Journal of Biological Chemistry. 2003;278(20):18256.
152. Pietschmann T, Kaul A, Koutsoudakis G, Shavinskaya A, Kallis S, Steinmann E, et al., editors. Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. 2006: National Acad Sciences.
153. Ma Y, Anantpadma M, Timpe JM, Shanmugam S, Singh SM, Lemon SM, et al. Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins. Journal of virology. 2011;85(1):86-97.
154. Jirasko V, Montserret R, Appel N, Janvier A, Eustachi L, Brohm C, et al. Structural and functional characterization of nonstructural protein 2 for its role in hepatitis C virus assembly. Journal of Biological Chemistry. 2008;283(42):28546-62.

155. Reed K, Rice C. Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. Current topics in microbiology and immunology. 2000;242:55.
156. Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM. A second hepatitis C virus-encoded proteinase. Proceedings of the National Academy of Sciences. 1993;90(22):10583.
157. Hijikata M, Mizushima H, Tanji Y, Komoda Y, Hirowatari Y, Akagi T, et al. Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus. Proceedings of the National Academy of Sciences. 1993;90(22):10773-7.
158. Lorenz IC, Marcotrigiano J, Dentzer TG, Rice CM. Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. Nature. 2006;442(7104):831-5.
159. Pallaoro M, Lahm A, Biasiol G, Brunetti M, Nardella C, Orsatti L, et al. Characterization of the hepatitis C virus NS2/3 processing reaction by using a purified precursor protein. Journal of virology. 2001;75(20):9939.
160. Welbourn S, Pause A. The hepatitis C virus NS2/3 protease. Current issues in molecular biology. 2007;9(1):63.
161. Darke PL, Jacobs AR, Waxman L, Kuo LC. Inhibition of hepatitis C virus NS2/3 processing by NS4A peptides. Journal of Biological Chemistry. 1999;274(49):34511-4.
162. Kolykhalov AA, Mihalik K, Feinstone SM, Rice CM. Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3'nontranslated region are essential for virus replication in vivo. Journal of virology. 2000;74(4):2046.
163. Lohmann V, Körner F, Koch JO, Herian U, Theilmann L, Bartenschlager R. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. Science. 1999;285(5424):110.
164. Lackner T, Muller A, Pankraz A, Becher P, Thiel HJ, Gorbalenya A, et al. Temporal modulation of an autoprotease is crucial for replication and pathogenicity of an RNA virus. Journal of virology. 2004;78(19):10765.
165. Agapov EV, Murray CL, Frolov I, Qu L, Myers TM, Rice CM. Uncleaved NS2-3 is required for production of infectious bovine viral diarrhea virus. Journal of virology. 2004;78(5):2414.
166. Suzich J, Tamura J, Palmer-Hill F, Warrener P, Grakoui A, Rice C, et al. Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. Journal of virology. 1993;67(10):6152-8.
167. Zhang C, Cai Z, Kim Y-C, Kumar R, Yuan F, Shi P-Y, et al. Stimulation of hepatitis C virus (HCV) nonstructural protein 3 (NS3) helicase activity by the NS3 protease domain and by HCV RNA-dependent RNA polymerase. Journal of virology. 2005;79(14):8687-97.
168. Gallinari P, Paolini C, Brennan D, Nardi C, Steinkühler C, De Francesco R. Modulation of hepatitis C virus NS3 protease and helicase activities through the interaction with NS4A. Biochemistry. 1999;38(17):5620-32.
169. Kuang W-F, Lin Y-C, Jean F, Huang Y-W, Tai C-L, Chen D-S, et al. Hepatitis C virus NS3 RNA helicase activity is modulated by the two domains of NS3 and NS4A. Biochemical and biophysical research communications. 2004;317(1):211-7.
170. Schregel V, Jacobi S, Penin F, Tautz N. Hepatitis C virus NS2 is a protease stimulated by cofactor domains in NS3. Proceedings of the National Academy of Sciences. 2009;106(13):5342-7.
171. Kim J, Morgenstern K, Lin C, Fox T, Dwyer M, Landro J, et al. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell. 1996;87(2):343-55.
172. Archer SJ, Camac DM, Wu ZJ, Farrow NA, Domaille PJ, Wasserman ZR, et al. Hepatitis C virus NS3 protease requires its NS4A cofactor peptide for optimal binding of a boronic acid inhibitor as shown by NMR. Chemistry & biology. 2002;9(1):79-92.
173. Major ME, Feinstone SM. The molecular virology of hepatitis C. Hepatology. 1997;25(6):1527-38.
174. Li K, Foy E, Ferreon JC, Nakamura M, Ferreon AC, Ikeda M, et al. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF.

Proceedings of the National Academy of Sciences of the United States of America. 2005;102(8):2992-7.
175. Li X-D, Sun L, Seth RB, Pineda G, Chen ZJ. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(49):17717-22.
176. Lamarre D, Anderson PC, Bailey M, Beaulieu P, Bolger G, Bonneau P, et al. An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. Nature. 2003;426(6963):186-9.
177. Kwo PY, Lawitz EJ, McCone J, Schiff ER, Vierling JM, Pound D, et al. Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): an open-label, randomised, multicentre phase 2 trial. The Lancet. 2010;376(9742):705-16.
178. Marks KM, Jacobson IM. The first wave: HCV NS3 protease inhibitors telaprevir and boceprevir. Antiviral therapy. 2012;17(6):1119.
179. Bartels DJ, Zhou Y, Zhang EZ, Marcial M, Byrn RA, Pfeiffer T, et al. Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3· 4A protease inhibitors in treatment-naive subjects. Journal of Infectious Diseases. 2008;198(6):800-7.
180. Preugschat F, Averett DR, Clarke BE, Porter DJ. A steady-state and pre-steady-state kinetic analysis of the NTPase activity associated with the hepatitis C virus NS3 helicase domain. Journal of Biological Chemistry. 1996;271(40):24449-57.
181. Tai C-L, Chi W-K, Chen D-S, Hwang L-H. The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). Journal of virology. 1996;70(12):8477-84.
182. Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, Murcko MA, et al. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure. 1998;6(1):89-100.
183. Myong S, Bruno MM, Pyle AM, Ha T. Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. Science. 2007;317(5837):513-6.
184. Gwack Y, Kim DW, Han JH, Choe J. Characterization of RNA binding activity and RNA helicase activity of the hepatitis C virus NS3 protein. Biochemical and biophysical research communications. 1996;225(2):654-9.
185. Lindenbach BD, Rice CM. The ins and outs of hepatitis C virus entry and assembly. Nature Reviews Microbiology. 2013;11(10):688-700.
186. Morikawa K, Lange C, Gouttenoire J, Meylan E, Brass V, Penin F, et al. Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus. Journal of viral hepatitis. 2011;18(5):305-15.
187. Tanji Y, Hijikata M, Satoh S, Kaneko T, Shimotohno K. Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. Journal of virology. 1995;69(3):1575-81.
188. Horner SM, Liu HM, Park HS, Briley J, Gale M. Mitochondrial-associated endoplasmic reticulum membranes (MAM) form innate immune synapses and are targeted by hepatitis C virus. Proceedings of the National Academy of Sciences. 2011;108(35):14590-5.
189. Phan T, Kohlway A, Dimberu P, Pyle AM, Lindenbach BD. The acidic domain of hepatitis C virus NS4A contributes to RNA replication and virus particle assembly. Journal of virology. 2011;85(3):1193-204.
190. Hügle T, Fehrmann F, Bieck E, Kohara M, Kräusslich HG, Rice CM, et al. The hepatitis C virus nonstructural protein 4B is an integral endoplasmic reticulum membrane protein. Virology. 2001;284(1):70-81.
191. Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. Kinetic and structural analyses of hepatitis C virus polyprotein processing. Journal of virology. 1994;68(8):5045.
192. Pietschmann T, Lohmann V, Rutter G, Kurpanek K, Bartenschlager R. Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. Journal of virology. 2001;75(3):1252.

193. Brass V, Berke JM, Montserret R, Blum HE, Penin F, Moradpour D. Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex. Proceedings of the National Academy of Sciences. 2008;105(38):14545.
194. Lundin M, Monne M, Widell A, Von Heijne G, Persson MAA. Topology of the membrane-associated hepatitis C virus protein NS4B. Journal of virology. 2003;77(9):5428.
195. Egger D, Wolk B, Gosert R, Bianchi L, Blum HE, Moradpour D, et al. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. Journal of virology. 2002;76(12):5974.
196. Gosert R, Egger D, Lohmann V, Bartenschlager R, Blum HE, Bienz K, et al. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. Journal of virology. 2003;77(9):5487.
197. Gouttenoire J, Castet V, Montserret R, Arora N, Raussens V, Ruysschaert JM, et al. Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B. Journal of virology. 2009;83(12):6257.
198. Gouttenoire J, Roingeard P, Penin F, Moradpour D. Amphipathic {alpha}-Helix AH2 Is a Major Determinant for the Oligomerization of Hepatitis C Virus Nonstructural Protein 4B. Journal of virology. 2010;84(24):12529.
199. Gouttenoire J, Penin F, Moradpour D. Hepatitis C virus nonstructural protein 4B: a journey into unexplored territory. Reviews in Medical Virology. 2010;20(2):117-29.
200. Lundin M, Lindström H, Grönwall C, Persson MAA. Dual topology of the processed hepatitis C virus protein NS4B is influenced by the NS5A protein. Journal of general virology. 2006;87(11):3263.
201. Liefhebber JMP, Brandt BW, Broer R, Spaan WJM, Van Leeuwen HC. Hepatitis C virus NS4B carboxy terminal domain is a membrane binding domain. Virology journal. 2009;6(1):62.
202. Paul D, Romero-Brey I, Gouttenoire J, Stoitsova S, Krijnse-Locker J, Moradpour D, et al. NS4B Self-Interaction through Conserved C-Terminal Elements Is Required for the Establishment of Functional Hepatitis C Virus Replication Complexes. Journal of virology. 2011;85(14):6963.
203. Thompson AA, Zou A, Yan J, Duggal R, Hao W, Molina D, et al. Biochemical Characterization of Recombinant Hepatitis C Virus Nonstructural Protein 4B: Evidence for ATP/GTP Hydrolysis and Adenylate Kinase Activity†. Biochemistry. 2009;48(5):906-16.
204. Jones DM, Patel AH, Targett-Adams P, McLauchlan J. The hepatitis C virus NS4B protein can trans-complement viral RNA replication and modulates production of infectious virus. Journal of virology. 2009;83(5):2163.
205. Einav S, Elazar M, Danieli T, Glenn JS. A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication. Journal of virology. 2004;78(20):11288.
206. Tasaka M, Sakamoto N, Itakura Y, Nakagawa M, Itsui Y, Sekine-Osajima Y, et al. Hepatitis C virus non-structural proteins responsible for suppression of the RIG-I/Cardif-induced interferon response. Journal of general virology. 2007;88(12):3323-33.
207. Xu J, Liu S, Xu Y, Tien P, Gao G. Identification of the nonstructural protein 4B of hepatitis C virus as a factor that inhibits the antiviral activity of interferon-alpha. Virus research. 2009;141(1):55-62.
208. Tripathi LP, Kataoka C, Taguwa S, Moriishi K, Mori Y, Matsuura Y, et al. Network based analysis of hepatitis C virus Core and NS4B protein interactions. Mol BioSyst. 2010;6(12):2539-53.
209. Ross-Thriepland D, Harris M. Hepatitis C virus NS5A: enigmatic but still promiscuous 10 years on! Journal of General Virology. 2015;96(Pt 4):727-38.
210. McPhee F, Hernandez D, Yu F, Ueland J, Monikowski A, Carifa A, et al. Resistance analysis of hepatitis C virus genotype 1 prior treatment null responders receiving daclatasvir and asunaprevir. Hepatology. 2013;58(3):902-11.
211. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. New England Journal of Medicine. 1996;334(2):77-82.
212. Kato N. Molecular Virology of Hepatitis Virus. Acta Med Okayama. 2001;55:133-59.

213. Moradpour D, Penin F. Hepatitis C virus proteins: from structure to function. Hepatitis C Virus: From Molecular Virology to Antiviral Therapy: Springer; 2013. p. 113-42.
214. Koff R. Review article: The efficacy and safety of sofosbuvir, a novel, oral nucleotide NS5B polymerase inhibitor, in the treatment of chronic hepatitis C virus infection. Alimentary Pharmacology & Therapeutics. 2014;39(5):478-87.
215. Jiang J, Wu X, Tang H, Luo G. Apolipoprotein E mediates attachment of clinical hepatitis C virus to hepatocytes by binding to cell surface heparan sulfate proteoglycan receptors. PloS one. 2013;8(7):e67982.
216. Bartosch B, Dubuisson J, Cosset FL. Infectious hepatitis C virus pseudo-particles containing functional E1–E2 envelope protein complexes. The Journal of experimental medicine. 2003;197(5):633.
217. Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. Nature Reviews Microbiology. 2007;5(6):453-63.
218. Cormier EG, Durso RJ, Tsamis F, Boussemart L, Manix C, Olson WC, et al. L-SIGN (CD209L) and DC-SIGN (CD209) mediate transinfection of liver cells by hepatitis C virus. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(39):14067.
219. Gardner JP, Durso RJ, Arrigale RR, Donovan GP, Maddon PJ, Dragic T, et al. L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(8):4498.
220. Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, et al. Binding of hepatitis C virus to CD81. Science. 1998;282(5390):938.
221. Scarselli E, Ansuini H, Cerino R, Roccasecca RM, Acali S, Filocamo G, et al. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. The EMBO journal. 2002;21(19):5017-25.
222. Bartosch B, Vitelli A, Granier C, Goujon C, Dubuisson J, Pascale S, et al. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. Journal of Biological Chemistry. 2003;278(43):41624.
223. Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, De Jong YP, et al. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. Nature. 2009;457(7231):882-6.
224. Evans MJ, Von Hahn T, Tscherne DM, Syder AJ, Panis M, Wölk B, et al. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature. 2007;446(7137):801-5.
225. Harris HJ, Davis C, Mullins JG, Hu K, Goodall M, Farquhar MJ, et al. Claudin association with CD81 defines hepatitis C virus entry. Journal of Biological Chemistry. 2010;285(27):21092-102.
226. Coller KE, Berger KL, Heaton NS, Cooper JD, Yoon R, Randall G. RNA interference and single particle tracking analysis of hepatitis C virus endocytosis. 2009.
227. Zona L, Lupberger J, Sidahmed-Adrar N, Thumann C, Harris HJ, Barnes A, et al. HRas signal transduction promotes hepatitis C virus cell entry by triggering assembly of the host tetraspanin receptor complex. Cell host & microbe. 2013;13(3):302-13.
228. Farquhar MJ, Hu K, Harris HJ, Davis C, Brimacombe CL, Fletcher SJ, et al. Hepatitis C virus induces CD81 and claudin-1 endocytosis. Journal of virology. 2012;86(8):4305-16.
229. Blanchard E, Belouzard S, Goueslain L, Wakita T, Dubuisson J, Wychowski C, et al. Hepatitis C virus entry depends on clathrin-mediated endocytosis. Journal of virology. 2006;80(14):6964.
230. Meertens L, Bertaux C, Dragic T. Hepatitis C virus entry requires a critical postinternalization step and delivery to early endosomes via clathrin-coated vesicles. Journal of virology. 2006;80(23):11571.
231. Hsu M, Zhang J, Flint M, Logvinoff C, Cheng-Mayer C, Rice CM, et al. Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. Proceedings of the National Academy of Sciences. 2003;100(12):7271.
232. Sharma NR, Mateu G, Dreux M, Grakoui A, Cosset F-L, Melikyan GB. Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion. Journal of Biological Chemistry. 2011;286(35):30361-76.

233. Liu S, Yang W, Shen L, Turner JR, Coyne CB, Wang T. Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. Journal of virology. 2009;83(4):2011.
234. Mee CJ, Grove J, Harris HJ, Hu K, Balfe P, McKeating JA. Effect of cell polarization on hepatitis C virus entry. Journal of virology. 2008;82(1):461.
235. Mee CJ, Harris HJ, Farquhar MJ, Wilson G, Reynolds G, Davis C, et al. Polarization restricts hepatitis C virus entry into HepG2 hepatoma cells. Journal of virology. 2009;83(12):6211.
236. Zheng A, Yuan F, Li Y, Zhu F, Hou P, Li J, et al. Claudin-6 and claudin-9 function as additional coreceptors for hepatitis C virus. Journal of virology. 2007;81(22):12465.
237. Lindenbach BD, Rice CM. Unravelling hepatitis C virus replication from genome to function. Nature. 2005;436(7053):933-8.
238. Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CUT. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initation of hepatitis C and classical swine fever virus RNAs. Genes & development. 1998;12(1):67.
239. Siridechadilok B, Fraser CS, Hall RJ, Doudna JA, Nogales E. Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science. 2005;310(5753):1513.
240. George A, Panda S, Kudmulwar D, Chhatbar SP, Nayak SC, Krishnan HH. Hepatitis C Virus NS5A Binds to the mRNA Cap-binding Eukaryotic Translation Initiation 4F (eIF4F) Complex and Up-regulates Host Translation Initiation Machinery through eIF4E-binding Protein 1 Inactivation. Journal of Biological Chemistry. 2012;287(7):5042-58.
241. Henke JI, Goergen D, Zheng J, Song Y, Schüttler CG, Fehr C, et al. microRNA-122 stimulates translation of hepatitis C virus RNA. The EMBO journal. 2008;27(24):3300-10.
242. Reiss S, Rebhan I, Backes P, Romero-Brey I, Erfle H, Matula P, et al. Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. Cell host & microbe. 2011;9(1):32-45.
243. Kapadia SB, Chisari FV. Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(7):2561.
244. Bader T, Fazili J, Madhoun M, Aston C, Hughes D, Rizvi S, et al. Fluvastatin inhibits hepatitis C replication in humans. The American journal of gastroenterology. 2008;103(6):1383-9.
245. Li Y, Masaki T, Yamane D, McGivern DR, Lemon SM. Competing and noncompeting activities of miR-122 and the 5′ exonuclease Xrn1 in regulation of hepatitis C virus replication. Proceedings of the National Academy of Sciences. 2013;110(5):1881-6.
246. Diamond DL, Syder AJ, Jacobs JM, Sorensen CM, Walters K-A, Proll SC, et al. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. PLoS Pathog. 2010;6(1):e1000719.
247. Targett-Adams P, Boulant S, McLauchlan J. Visualization of double-stranded RNA in cells supporting hepatitis C virus RNA replication. Journal of virology. 2008;82(5):2182-95.
248. Saeed M, Andreo U, Chung H-Y, Espiritu C, Branch AD, Silva JM, et al. SEC14L2 enables pan-genotype HCV replication in cell culture. Nature. 2015.
249. Reiss S, Harak C, Romero-Brey I, Radujkovic D, Klein R, Ruggieri A, et al. The lipid kinase phosphatidylinositol-4 kinase III alpha regulates the phosphorylation status of hepatitis C virus NS5A. 2013.
250. Paul D, Hoppe S, Saher G, Krijnse-Locker J, Bartenschlager R. Morphological and biochemical characterization of the membranous hepatitis C virus replication compartment. Journal of virology. 2013;87(19):10612-27.
251. Boulant S, Vanbelle C, Ebel C, Penin F, Lavergne J-P. Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. Journal of virology. 2005;79(17):11353-65.

252. Barba G, Harper F, Harada T, Kohara M, Goulinet S, Matsuura Y, et al. Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. Proceedings of the National Academy of Sciences. 1997;94(4):1200-5.
253. Boulant S, Douglas MW, Moody L, Budkowska A, Targett-Adams P, McLauchlan J. Hepatitis C Virus Core Protein Induces Lipid Droplet Redistribution in a Microtubule-and Dynein-Dependent Manner. Traffic. 2008;9(8):1268-82.
254. Neveu G, Barouch-Bentov R, Ziv-Av A, Gerber D, Jacob Y, Einav S. Identification and targeting of an interaction between a tyrosine motif within hepatitis C virus core protein and AP2M1 essential for viral assembly. PLoS Pathog. 2012;8(8):e1002845-e.
255. Boson B, Granio O, Bartenschlager R, Cosset F-L. A concerted action of hepatitis C virus p7 and nonstructural protein 2 regulates core localization at the endoplasmic reticulum and virus assembly. PLoS Pathog. 2011;7(7):e1002144-e.
256. Dubuisson J, Hsu HH, Cheung RC, Greenberg HB, Russell DG, Rice CM. Formation and intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses. Journal of virology. 1994;68(10):6147-60.
257. Stapleford KA, Lindenbach BD. Hepatitis C virus NS2 coordinates virus particle assembly through physical interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. Journal of virology. 2011;85(4):1706-17.
258. Gentzsch J, Brohm C, Steinmann E, Friesland M, Menzel N, Vieyres G, et al. Hepatitis C virus p7 is critical for capsid assembly and envelopment. 2013.
259. Ma Y, Anantpadma M, Timpe JM, Shanmugam S, Singh SM, Lemon SM, et al. Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins. Journal of virology. 2011;85(1):86-97.
260. Tellinghuisen TL, Foss KL, Treadaway J. Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. 2008.
261. Lindenbach BD. Virion assembly and release. Hepatitis C Virus: From Molecular Virology to Antiviral Therapy: Springer; 2013. p. 199-218.
262. Huang H, Sun F, Owen DM, Li W, Chen Y, Gale M, et al. Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins. Proceedings of the National Academy of Sciences. 2007;104(14):5848-53.
263. Hueging K, Doepke M, Vieyres G, Bankwitz D, Frentzen A, Doerrbecker J, et al. Apolipoprotein E codetermines tissue tropism of hepatitis C virus and is crucial for viral cell-to-cell transmission by contributing to a postenvelopment step of assembly. Journal of virology. 2014;88(3):1433-46.
264. Saito T, Owen DM, Jiang F, Marcotrigiano J, Gale Jr M. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature. 2008;454(7203):523-7.
265. Thimme R, Binder M, Bartenschlager R. Failure of innate and adaptive immune responses in controlling hepatitis C virus infection. FEMS microbiology reviews. 2012;36(3):663-83.
266. Saito T, Hirai R, Loo Y-M, Owen D, Johnson CL, Sinha SC, et al. Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. Proceedings of the National Academy of Sciences. 2007;104(2):582-7.
267. Arnaud N, Dabo S, Akazawa D, Fukasawa M, Shinkai-Ouchi F, Hugon J, et al. Hepatitis C virus reveals a novel early control in acute immune response. PLoS Pathog. 2011;7(10):e1002289.
268. You S, Rice CM. 3′ RNA elements in hepatitis C virus replication: kissing partners and long poly (U). Journal of virology. 2008;82(1):184-95.
269. Chen L, Borozan I, Feld J, Sun J, Tannis L-L, Coltescu C, et al. Hepatic gene expression discriminates responders and nonresponders in treatment of chronic hepatitis C viral infection. Gastroenterology. 2005;128(5):1437-44.
270. Binder M, Eberle F, Seitz S, Mücke N, Hüber CM, Kiani N, et al. Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid-inducible gene-I (RIG-I). Journal of Biological Chemistry. 2011;286(31):27278-87.

271. Baril M, Racine M-E, Penin F, Lamarre D. MAVS dimer is a crucial signaling component of innate immunity and the target of hepatitis C virus NS3/4A protease. Journal of virology. 2009;83(3):1299-311.
272. Chang S, Dolganiuc A, Szabo G. Toll-like receptors 1 and 6 are involved in TLR2-mediated macrophage activation by hepatitis C virus core and NS3 proteins. Journal of leukocyte biology. 2007;82(3):479-87.
273. Dolganiuc A, Kodys K, Kopasz A, Marshall C, Do T, Romics L, et al. Hepatitis C virus core and nonstructural protein 3 proteins induce pro-and anti-inflammatory cytokines and inhibit dendritic cell differentiation. The Journal of Immunology. 2003;170(11):5615-24.
274. Li K, Li NL, Wei D, Pfeffer SR, Fan M, Pfeffer LM. Activation of chemokine and inflammatory cytokine response in hepatitis C virus–infected hepatocytes depends on toll-like receptor 3 sensing of hepatitis C virus double-stranded RNA intermediates. Hepatology. 2012;55(3):666-75.
275. Wang N, Liang Y, Devaraj S, Wang J, Lemon SM, Li K. Toll-like receptor 3 mediates establishment of an antiviral state against hepatitis C virus in hepatoma cells. Journal of virology. 2009;83(19):9824-34.
276. Negash AA, Ramos HJ, Crochet N, Lau D, Doehle B, Papic N, et al. IL-1beta production through the NLRP3 inflammasome by hepatic macrophages links hepatitis C virus infection with liver inflammation and disease. PLoS Pathog. 2013;9(4):e1003330.
277. Burdette D, Haskett A, Presser L, McRae S, Iqbal J, Waris G. Hepatitis C virus activates interleukin-1β via caspase-1-inflammasome complex. Journal of General Virology. 2012;93(2):235-46.
278. Shrivastava S, Mukherjee A, Ray R, Ray RB. Hepatitis C virus induces interleukin-1β (IL-1β)/IL-18 in circulatory and resident liver macrophages. Journal of virology. 2013;87(22):12284-90.
279. Thimme R, Bukh J, Spangenberg HC, Wieland S, Pemberton J, Steiger C, et al. Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. Proceedings of the National Academy of Sciences. 2002;99(24):15661-8.
280. Pestka JM, Zeisel MB, Bläser E, Schürmann P, Bartosch B, Cosset F-L, et al. Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. Proceedings of the National Academy of Sciences. 2007;104(14):6025-30.
281. Christie J, Healey C, Watson J, Wong V, Duddridge M, Snowden N, et al. Clinical outcome of hypogammaglobulinaemic patients following outbreak of acute hepatitis C: 2 year follow up. Clinical and experimental immunology. 1997;110(1):4.
282. Palmer BA, Moreau I, Levis J, Harty C, Crosbie O, Kenny-Walsh E, et al. Insertion and recombination events at hypervariable region 1 over 9.6 years of hepatitis C virus chronic infection. Journal of General Virology. 2012;93(Pt 12):2614-24.
283. Moreau I, O'Sullivan H, Murray C, Levis J, Crosbie O, Kenny-Walsh E, et al. Separation of Hepatitis C genotype 4a into IgG-depleted and IgG-enriched fractions reveals a unique quasispecies profile. Virology journal. 2008;5(1):103.
284. Skums P, Bunimovich L, Khudyakov Y. Antigenic cooperation among intrahost HCV variants organized into a complex network of cross-immunoreactivity. Proceedings of the National Academy of Sciences. 2015;112(21):6653-8.
285. Cooper S, Erickson AL, Adams EJ, Kansopon J, Weiner AJ, Chien DY, et al. Analysis of a successful immune response against hepatitis C virus. Immunity. 1999;10(4):439-49.
286. Shoukry NH, Grakoui A, Houghton M, Chien DY, Ghrayeb J, Reimann KA, et al. Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. The Journal of experimental medicine. 2003;197(12):1645-55.
287. Grakoui A, Shoukry NH, Woollard DJ, Han J-H, Hanson HL, Ghrayeb J, et al. HCV persistence and immune evasion in the absence of memory T cell help. Science. 2003;302(5645):659-62.
288. Dazert E, Neumann-Haefelin C, Bressanelli S, Fitzmaurice K, Kort J, Timm J, et al. Loss of viral fitness and cross-recognition by CD8+ T cells limit HCV escape from a protective HLA-B27–restricted human immune response. The Journal of clinical investigation. 2009;119(2):376.

289. McKiernan SM, Hagan R, Curry M, McDonald GS, Kelly A, Nolan N, et al. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. Hepatology. 2004;40(1):108-14.
290. Nakamoto N, Cho H, Shaked A, Olthoff K, Valiga ME, Kaminski M, et al. Synergistic reversal of intrahepatic HCV-specific CD8 T cell exhaustion by combined PD-1/CTLA-4 blockade. PLoS Pathog. 2009;5(2):e1000313.
291. Drake JW, Holland JJ. Mutation rates among RNA viruses. Proceedings of the National Academy of Sciences. 1999;96(24):13910-3.
292. Cuevas JM, González-Candelas F, Moya A, Sanjuán R. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. Journal of virology. 2009;83(11):5760-4.
293. Eigen M. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften. 1971;58(10):465-523.
294. Domingo E, Martínez-Salas E, Sobrino F, de la Torre JC, Portela A, Ortín J, et al. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene. 1985;40(1):1-8.
295. Biebricher CK, Eigen M. The error threshold. Virus research. 2005;107(2):117-27.
296. Bull JJ, Sanjuan R, Wilke CO. Theory of lethal mutagenesis for viruses. Journal of virology. 2007;81(6):2930-9.
297. Gaudieri S, Rauch A, Pfafferott K, Barnes E, Cheng W, McCaughan G, et al. Hepatitis C virus drug resistance and immune-driven adaptations: Relevance to new antiviral therapy. Hepatology. 2009;49(4):1069-82.
298. Chao L. Fitness of RNA virus decreased by Muller's ratchet. 1990.
299. Duarte E, Clarke D, Moya A, Domingo E, Holland J. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. Proceedings of the National Academy of Sciences. 1992;89(13):6015-9.
300. Novella IS, Duarte EA, Elena SF, Moya A, Domingo E, Holland JJ. Exponential increases of RNA virus fitness during large population transmissions. Proceedings of the National Academy of Sciences. 1995;92(13):5841-4.
301. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. PLoS Pathog. 2011;7(9):e1002243.
302. Schmidt-Martin D, Crosbie O, Kenny-Walsh E, Fanning LJ. Intensive temporal mapping of Hepatitis C Hypervariable Region 1 quasispecies provides novel insights into HCV evolution in chronic infection. Journal of General Virology. 2015:vir. 0.000149.
303. Herring BL, Tsui R, Peddada L, Busch M, Delwart EL. Wide range of quasispecies diversity during primary hepatitis C virus infection. Journal of virology. 2005;79(7):4340-6.
304. Thomson EC, Smith JA, Klenerman P. The natural history of early hepatitis C virus evolution; lessons from a global outbreak in human immunodeficiency virus-1-infected individuals. Journal of General Virology. 2011;92(10):2227-36.
305. Wang W, Lin J, Tan D, Xu Y, Brunt EM, Fan X, et al. Divergent quasispecies evolution in de novo hepatitis C virus infection associated with bone marrow transplantation. Biochemical and biophysical research communications. 2011;414(1):148-52.
306. Manzin A, Solforosi L, Petrelli E, Macarri G, Tosone G, Piazza M, et al. Evolution of hypervariable region 1 of hepatitis C virus in primary infection. Journal of virology. 1998;72(7):6271-6.
307. Dowd KA, Netski DM, Wang XH, Cox AL, Ray SC. Selection pressure from neutralizing antibodies drives sequence evolution during acute infection with hepatitis C virus. Gastroenterology. 2009;136(7):2377-86.
308. Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, et al. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science. 2000;288(5464):339-44.

309. Ramachandran S, Campo DS, Dimitrova ZE, Xia G-l, Purdy MA, Khudyakov YE. Temporal variations in the hepatitis C virus intrahost population during chronic infection. Journal of virology. 2011;85(13):6369-80.
310. Liu L, Fisher BE, Dowd KA, Astemborski J, Cox AL, Ray SC. Acceleration of hepatitis C virus envelope evolution in humans is consistent with progressive humoral immune selection during the transition from acute to chronic infection. Journal of virology. 2010;84(10):5067-77.
311. Guglietta S, Garbuglia AR, Salichos L, Ruggeri L, Folgori A, Perrone MP, et al. Impact of viral selected mutations on T cell mediated immunity in chronically evolving and self limiting acute HCV infection. Virology. 2009;386(2):398-406.
312. Scottà C, Garbuglia AR, Ruggeri L, Spada E, Laurenti L, Perrone MP, et al. Influence of specific CD4+ T cells and antibodies on evolution of hypervariable region 1 during acute HCV infection. Journal of hepatology. 2008;48(2):216-28.
313. Ray SC, Wang Y-M, Laeyendecker O, Ticehurst JR, Villano SA, Thomas DL. Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. Journal of virology. 1999;73(4):2938-46.
314. Brown RJ, Juttla VS, Tarr AW, Finnis R, Irving WL, Hemsley S, et al. Evolutionary dynamics of hepatitis C virus envelope genes during chronic infection. Journal of General Virology. 2005;86(7):1931-42.
315. Cabot B, Martell Ma, Esteban JI, Piron M, Otero T, Esteban R, et al. Longitudinal evaluation of the structure of replicating and circulating hepatitis C virus quasispecies in nonprogressive chronic hepatitis C patients. Journal of virology. 2001;75(24):12005-13.
316. Ray SC, Fanning L, Wang X-H, Netski DM, Kenny-Walsh E, Thomas DL. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. The Journal of experimental medicine. 2005;201(11):1753-9.
317. Wang X-H, Netski DM, Astemborski J, Mehta SH, Torbenson MS, Thomas DL, et al. Progression of fibrosis during chronic hepatitis C is associated with rapid virus evolution. Journal of virology. 2007;81(12):6513-22.
318. Kao J-H, Chen P-J, Lai M-Y, Wang T-H, Chen D-S. Quasispecies of hepatitis C virus and genetic drift of the hypervariable region in chronic type C hepatitis. Journal of Infectious Diseases. 1995;172(1):261-4.
319. Qin H, Shire NJ, Keenan ED, Rouster SD, Eyster ME, Goedert JJ, et al. HCV quasispecies evolution: association with progression to end-stage liver disease in hemophiliacs infected with HCV or HCV/HIV. Blood. 2004.
320. Li H, McMahon BJ, McArdle S, Bruden D, Sullivan DG, Shelton D, et al. Hepatitis C virus envelope glycoprotein co-evolutionary dynamics during chronic hepatitis C. Virology. 2008;375(2):580-91.
321. Kuntzen T, Timm J, Berical A, Lewis-Ximenez LL, Jones A, Nolan B, et al. Viral sequence evolution in acute hepatitis C virus infection. Journal of virology. 2007;81(21):11658-68.
322. Lyra AC, Fan X, Lang DM, Yusim K, Ramrakhiani S, Brunt EM, et al. Evolution of hepatitis C viral quasispecies after liver transplantation. Gastroenterology. 2002;123(5):1485-93.
323. Gretch DR, Polyak SJ, Wilson JJ, Carithers R, Perkins JD, Corey L. Tracking hepatitis C virus quasispecies major and minor variants in symptomatic and asymptomatic liver transplant recipients. Journal of virology. 1996;70(11):7622-31.
324. Garcia JM, del Campo Terron S, Garcia GM, Gonzalez MG, de Vicente Lopez E, Vazquez-Garza JN, et al., editors. Analysis of hepatitis C viral quasispecies in liver transplantation. Transplantation proceedings; 2003: Elsevier.
325. Li H, Sullivan DG, Feuerborn N, McArdle S, Bekele K, Pal S, et al. Genetic diversity of hepatitis C virus predicts recurrent disease after liver transplantation. Virology. 2010;402(2):248-55.
326. Pessoa MG, Bzowej N, Berenguer M, Phung Y, Kim M, Ferrell L, et al. Evolution of hepatitis C virus quasispecies in patients with severe cholestatic hepatitis after liver transplantation. Hepatology. 1999;30(6):1513-20.

327. Fan X, Di Bisceglie AM. Diversification of hypervariable region 1 of hepatitis C virus after liver transplantation. Journal of medical virology. 2003;70(2):212-8.
328. Neau D, Jouvencel AC, Legrand E, Trimoulet P, Galperine T, Chitty I, et al. Hepatitis C virus genetic variability in 52 human immunodeficiency virus-coinfected patients. Journal of medical virology. 2003;71(1):41-8.
329. Toyoda H, Fukuda H, Koyama Y, Takamatsu J, Saito H, Hayakawa T. Effect of immunosuppression on composition of quasispecies population of hepatitis C virus in patients with chronic hepatitis C coinfected with human immunodeficiency virus. Journal of hepatology. 1997;26(5):975-82.
330. Bernini F, Ebranati E, De Maddalena C, Shkjezi R, Milazzo L, Lo Presti A, et al. Within-host dynamics of the hepatitis C virus quasispecies population in HIV-1/HCV coinfected patients. PloS one. 2011;6(1):e16551.
331. Shuhart MC, Sullivan DG, Bekele K, Harrington RD, Kitahata MM, Mathisen TL, et al. HIV infection and antiretroviral therapy: effect on hepatitis C virus quasispecies variability. Journal of Infectious Diseases. 2006;193(9):1211-8.
332. Mao Q, Ray SC, Laeyendecker O, Ticehurst JR, Strathdee SA, Vlahov D, et al. Human immunodeficiency virus seroconversion and evolution of the hepatitis C virus quasispecies. Journal of virology. 2001;75(7):3259-67.
333. Netski DM, Mao Q, Ray SC, Klein RS. Genetic divergence of HCV: The role of HIV-related immunosuppression. Journal of acquired immune deficiency syndromes (1999). 2008;49(2):136.
334. Farci P, Strazzera R, Alter HJ, Farci S, Degioannis D, Coiana A, et al. Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. Proceedings of the National Academy of Sciences. 2002;99(5):3081-6.
335. Boulestin A, Sandres-Sauné K, Payen JL, Alric L, Dubois M, Pasquier C, et al. Genetic heterogeneity of the envelope 2 gene and eradication of hepatitis C virus after a second course of interferon-α. Journal of medical virology. 2002;68(2):221-8.
336. Frasca L, Del Porto P, Tuosto L, Marinari B, Scottà C, Carbonari M, et al. Hypervariable region 1 variants act as TCR antagonists for hepatitis C virus-specific CD4+ T cells. The Journal of Immunology. 1999;163(2):650-8.
337. Cox AL, Mosbruger T, Mao Q, Liu Z, Wang X-H, Yang H-C, et al. Cellular immune selection with hepatitis C virus persistence in humans. The Journal of experimental medicine. 2005;201(11):1741-52.
338. Kato N, Sekiya H, Ootsuyama Y, Nakazawa T, Hijikata M, Ohkoshi S, et al. Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. Journal of virology. 1993;67(7):3923-30.
339. Ray SC, Mao Q, Lanford RE, Bassett S, Laeyendecker O, Wang Y-M, et al. Hypervariable region 1 sequence stability during hepatitis C virus replication in chimpanzees. Journal of virology. 2000;74(7):3058-66.
340. Hézode C, Fontaine H, Dorival C, Larrey D, Zoulim F, Canva V, et al. Triple therapy in treatment-experienced patients with hcv-cirrhosis in a multicentre cohort of the french early access programme (anrs co20-cupic)-nct01514890. Journal of hepatology. 2013.
341. Okada SI, Akahane Y, Suzuki H, Okamoto H, Mishiro S. The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. Hepatology. 1992;16(3):619-24.
342. Nakazawa T, Kato N, Ohkoshi S, Shibuya A, Shimotohno K. Characterization of the 5′ noncoding and structural region of the hepatitis C virus genome from patients with non-A, non-B hepatitis responding differently to interferon treatment. Journal of hepatology. 1994;20(5):623-9.
343. Grahovac B, Bingulac-Popović J, Vucelić B, Hrstić I, Ostojić R, Dražić V, et al. Hypervariable region 1 of hepatitis C virus genome and response to interferon therapy. Clinical chemistry and laboratory medicine. 2000;38(9):905-10.

344. Koizumi K, Enomoto N, Kurosaki M, Murakami T, Izumi N, Marumo F, et al. Diversity of quasispecies in various disease stages of chronic hepatitis C virus infection and its significance in interferon treatment. Hepatology. 1995;22(1):30-5.
345. Moribe T, Hayashi N, Kanazawa Y, Mita E, Fusamoto H, Negi M, et al. Hepatitis C viral complexity detected by single-strand conformation polymorphism and response to interferon therapy. Gastroenterology. 1995;108(3):789-95.
346. Pawlotsky JM, Pellerin M, Bouvier M, Roudot-Thoraval F, Germanidis G, Bastie A, et al. Genetic complexity of the hypervariable region 1 (HVR1) of hepatitis C virus (HCV): influence on the characteristics of the infection and responses to interferon alfa therapy in patients with chronic hepatitis C. Journal of medical virology. 1998;54(4):256-64.
347. Yeh BI, Han KH, Lee HW, Sohn JH, Ryu WS, Yoon DJ, et al. Factors predictive of response to interferon-α therapy in hepatitis C virus type 1b infection. Journal of medical virology. 2002;66(4):481-7.
348. López-Labrador FX, Ampurdanès S, Giménez-Barcons M, Guilera M, Costa J, Jiménez de Anta MT, et al. Relationship of the genomic complexity of hepatitis C virus with liver disease severity and response to interferon in patients with chronic HCV genotype 1b interferon. Hepatology. 1999;29(3):897-903.
349. Sandres K, Dubois M, Pasquier C, Payen J, Alric L, Duffaut M, et al. Genetic heterogeneity of hypervariable region 1 of the hepatitis C virus (HCV) genome and sensitivity of HCV to alpha interferon therapy. Journal of virology. 2000;74(2):661-8.
350. Moreau I, Levis J, Crosbie O, Kenny-Walsh E, Fanning LJ. Correlation between pre-treatment quasispecies complexity and treatment outcome in chronic HCV genotype 3a. Virology journal. 2008;5(1):78.
351. Chambers TJ, Fan X, Droll DA, Hembrador E, Slater T, Nickells MW, et al. Quasispecies heterogeneity within the E1/E2 region as a pretreatment variable during pegylated interferon therapy of chronic hepatitis C virus infection. Journal of virology. 2005;79(5):3071-83.
352. Abbate I, Iacono OL, Di Stefano R, Cappiello G, Girardi E, Longo R, et al. HVR-1 quasispecies modifications occur early and are correlated to initial but not sustained response in HCV-infected patients treated with pegylated-or standard-interferon and ribavirin. Journal of hepatology. 2004;40(5):831-6.
353. Posada D. jModelTest: phylogenetic model averaging. Molecular biology and evolution. 2008;25(7):1253-6.
354. Fan X, Mao Q, Zhou D, Lu Y, Xing J, Xu Y, et al. High diversity of hepatitis C viral quasispecies is associated with early virological response in patients undergoing antiviral therapy. Hepatology. 2009;50(6):1765-72.
355. Watanabe H, Enomoto N, Nagayama K, Izumi N, Marumo F, Sato C, et al. Number and position of mutations in the interferon (IFN) sensitivity–determining region of the gene for nonstructural protein 5A correlate with IFN efficacy in hepatitis C virus genotype 1b infection. Journal of Infectious Diseases. 2001;183(8):1195-203.
356. Lusida MI, Nagano-Fujii M, Nidom CA, Handajani R, Fujita T, Oka K, et al. Correlation between mutations in the interferon sensitivity-determining region of NS5A protein and viral load of hepatitis C virus subtypes 1b, 1c, and 2a. Journal of clinical microbiology. 2001;39(11):3858-64.
357. Murakami T, Enomoto N, Kurosaki M, Izumi N, Marumo F, Sato C. Mutations in nonstructural protein 5A gene and response to interferon in hepatitis C virus genotype 2 infection. Hepatology. 1999;30(4):1045-53.
358. Shen C, Hu T, Shen L, Gao L, Xie W, Zhang J. Mutations in ISDR of NS5A gene influence interferon efficacy in Chinese patients with chronic hepatitis C virus genotype 1b infection. Journal of gastroenterology and hepatology. 2007;22(11):1898-903.
359. Zeuzem S, Lee J-H, Roth WK. Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. Hepatology. 1997;25(3):740-4.

360. Squadrito G, Orlando ME, Cacciola I, Rumi MG, Artini M, Picciotto A, et al. Long-term response to interferon alpha is unrelated to “interferon sensitivity determining region” variability in patients with chronic hepatitis C virus-1b infection. Journal of hepatology. 1999;30(6):1023-7.
361. Pascu M, Martus P, Höhne M, Wiedenmann B, Hopf U, Schreier E, et al. Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. Gut. 2004;53(9):1345-51.
362. Sáiz J-C, López-Labrador F-X, Ampurdanés S, Dopazo J, Forns X, Sánchez-Tapias J-M, et al. The prognostic relevance of the nonstructural 5A gene interferon sensitivity determining region is different in infections with genotype 1b and 3a isolates of hepatitis C virus. Journal of Infectious Diseases. 1998;177(4):839-47.
363. Frangeul L, Cresta P, Perrin M, Lunel F, Opolon P, Agut H, et al. Mutations in NS5A region of hepatitis C virus genome correlate with presence of NS5A antibodies and response to interferon therapy for most common European hepatitis C virus genotypes. Hepatology. 1998;28(6):1674-9.
364. Malta FdM, Medeiros-Filho JEMd, Azevedo RSd, Gonçalves L, Silva LCd, Carrilho FJ, et al. Sequencing of E2 and NS5A regions of HCV genotype 3a in Brazilian patients with chronic hepatitis. Memórias do Instituto Oswaldo Cruz. 2010;105(1):92-8.
365. Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, et al. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. Intervirology. 2005;48(6):372-80.
366. Akuta N, Suzuki F, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, et al. Amino acid substitutions in the hepatitis C virus core region are the important predictor of hepatocarcinogenesis. Hepatology. 2007;46(5):1357-64.
367. Nakamoto S, Imazeki F, Fukai K, Fujiwara K, Arai M, Kanda T, et al. Association between mutations in the core region of hepatitis C virus genotype 1 and hepatocellular carcinoma development. Journal of hepatology. 2010;52(1):72-8.
368. Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, et al. A matched case-controlled study of 48 and 72 weeks of peginterferon plus ribavirin combination therapy in patients infected with HCV genotype 1b in Japan: amino acid substitutions in HCV core region as predictor of sustained virological response. Journal of medical virology. 2009;81(3):452-8.
369. Funaoka Y, Sakamoto N, Suda G, Itsui Y, Nakagawa M, Kakinuma S, et al. Analysis of interferon signaling by infectious hepatitis C virus clones with substitutions of core amino acids 70 and 91. Journal of virology. 2011;85(12):5986-94.
370. Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, et al. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 2a high viral load and virological response to interferon-ribavirin combination therapy. Intervirology. 2009;52(6):301-9.
371. Hashimoto Y, Ochi H, Abe H, Hayashida Y, Tsuge M, Mitsui F, et al. Prediction of response to peginterferon-alfa-2b plus ribavirin therapy in Japanese patients infected with hepatitis C virus genotype 1b. Journal of medical virology. 2011;83(6):981-8.
372. Harvey CE, Post JJ, Palladinetti P, Freeman AJ, Ffrench RA, Kumar RK, et al. Expression of the chemokine IP-10 (CXCL10) by hepatocytes in chronic hepatitis C virus infection correlates with histological severity and lobular inflammation. Journal of leukocyte biology. 2003;74(3):360-9.
373. Diago M, Castellano G, Garcia-Samaniego J, Perez C, Fernandez I, Romero M, et al. Association of pretreatment serum interferon γ inducible protein 10 levels with sustained virological response to peginterferon plus ribavirin therapy in genotype 1 infected patients with chronic hepatitis C. Gut. 2006;55(3):374-9.
374. Lagging M, Romero AI, Westin J, Norkrans G, Dhillon AP, Pawlotsky JM, et al. IP-10 predicts viral response and therapeutic outcome in difficult-to-treat patients with HCV genotype 1 infection. Hepatology. 2006;44(6):1617-25.
375. Romero AI, Lagging M, Westin J, Dhillon AP, Dustin LB, Pawlotsky J-M, et al. Interferon (IFN)–γ–inducible protein–10: association with histological results, viral kinetics, and outcome during

treatment with pegylated IFN-α2a and ribavirin for chronic hepatitis C virus infection. Journal of Infectious Diseases. 2006;194(7):895-903.
376. Askarieh G, Alsiö Å, Pugnale P, Negro F, Ferrari C, Neumann AU, et al. Systemic and intrahepatic interferon-gamma-inducible protein 10 kDa predicts the first-phase decline in hepatitis C virus RNA and overall viral response to therapy in chronic hepatitis C. Hepatology. 2010;51(5):1523-30.
377. Falconer K, Askarieh G, Weis N, Hellstrand K, Alaeus A, Lagging M. IP-10 predicts the first phase decline of HCV RNA and overall viral response to therapy in patients co-infected with chronic hepatitis C virus infection and HIV. Scandinavian journal of infectious diseases. 2010;42(11-12):896-901.
378. Zeremski M, Dimova R, Astemborski J, Thomas DL, Talal AH. CXCL9 and CXCL10 chemokines as predictors of liver fibrosis in a cohort of primarily African-American injection drug users with chronic hepatitis C. Journal of Infectious Diseases. 2011;204(6):832-6.
379. Casrouge A, Decalf J, Ahloulay M, Lababidi C, Mansour H, Vallet-Pichard A, et al. Evidence for an antagonist form of the chemokine CXCL10 in patients chronically infected with HCV. The Journal of clinical investigation. 2011;121(1):308.
380. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature. 2009;461(7262):399-401.
381. Chen Y, Xu HX, Wang LJ, Liu XX, Mahato R, Zhao YR. Meta-analysis: IL28B polymorphisms predict sustained viral response in HCV patients treated with pegylated interferon-α and ribavirin. Alimentary Pharmacology & Therapeutics. 2012;36(2):91-103.
382. Moghaddam A, Melum E, Reinton N, Ring-Larsen H, Verbaan H, Bjøro K, et al. IL28B genetic variation and treatment response in patients with hepatitis C virus genotype 3 infection. Hepatology. 2011;53(3):746-54.
383. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'hUigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature. 2009;461(7265):798-801.
384. Noureddin M, Wright EC, Alter HJ, Clark S, Thomas E, Chen R, et al. Association of IL28B genotype with fibrosis progression and clinical outcomes in patients with chronic hepatitis C: A longitudinal analysis. Hepatology. 2013;58(5):1548-57.
385. Stättermayer AF, Stauber R, Hofer H, Rutter K, Beinhardt S, Scherzer TM, et al. Impact of IL28B genotype on the early and sustained virologic response in treatment-naive patients with chronic hepatitis C. Clinical Gastroenterology and Hepatology. 2011;9(4):344-50. e2.
386. Scherzer T-M, Hofer H, Staettermayer AF, Rutter K, Beinhardt S, Steindl-Munda P, et al. Early virologic response and IL28B polymorphisms in patients with chronic hepatitis C genotype 3 treated with peginterferon alfa-2a and ribavirin. Journal of hepatology. 2011;54(5):866-71.
387. Aziz H, Raza A, Ali K, Khattak JZK, Irfan J, Gill ML. Polymorphism of the IL28B gene (rs8099917, rs12979860) and virological response of Pakistani hepatitis C virus genotype 3 patients to pegylated interferon therapy. International Journal of Infectious Diseases. 2015;30:91-7.
388. Gupta AC, Trehanpati N, Sukriti S, Hissar S, Midha V, Sood A, et al. Interleukin-28b CC genotype predicts early treatment response and CT/TT genotypes predicts non-response in patients infected with HCV genotype 3. Journal of medical virology. 2014;86(4):707-12.
389. Sarrazin C, Susser S, Doehring A, Lange CM, Müller T, Schlecker C, et al. Importance of IL28B gene polymorphisms in hepatitis C virus genotype 2 and 3 infected patients. Journal of hepatology. 2011;54(3):415-21.
390. Rauch A, Kutalik Z, Descombes P, Cai T, Di Iulio J, Mueller T, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. Gastroenterology. 2010;138(4):1338-45. e7.
391. Van den Berg C, Grady B, Schinkel J, van de Laar T, Molenkamp R, van Houdt R, et al. Female sex and IL28B, a synergism for spontaneous viral clearance in hepatitis C virus (HCV) seroconverters from a community-based cohort. PloS one. 2011;6(11):e27555.

392. Sato M, Kondo M, Tateishi R, Fujiwara N, Kato N, Yoshida H, et al. Impact of IL28B genetic variation on HCV-induced liver fibrosis, inflammation, and steatosis: a meta-analysis. PloS one. 2014;9(3):e91822.
393. Lagging M, Askarieh G, Negro F, Bibert S, Soderholm J, Westin J, et al. Response prediction in chronic hepatitis C by assessment of IP-10 and IL28B-related single nucleotide polymorphisms. 2011.
394. Smith K, Suppiah V, O'Connor K, Berg T, Weltman M, Abate ML, et al. Identification of improved IL28B SNPs and haplotypes for prediction of drug response in treatment of hepatitis C using massively parallel sequencing in a cross-sectional European cohort: BioMed Central Limited; 2011.
395. Sulkowski M, Gardiner D, Rodriguez-Torres M, Reddy K, Hassanein T, Jacobson I, et al. 1417 Sustained Virologic Response With Daclatasvir Plus Sofosbuvir±Ribavirin (Rbv) In Chronic Hcv Genotype (Gt) 1-Infected Patients Who Previously Failed Telaprevir (Tvr) Or Boceprevir (Boc). Journal of hepatology. 2013;58:S570.
396. Palmer BA, Dimitrova Z, Skums P, Crosbie O, Kenny-Walsh E, Fanning LJ. Analysis of the evolution and structure of a complex intrahost viral population in chronic hepatitis C virus mapped by ultradeep pyrosequencing. Journal of virology. 2014;88(23):13709-21.
397. Skums P, Dimitrova Z, Campo DS, Vaughan G, Rossi L, Forbi JC, et al. Efficient error correction for next-generation sequencing of viral amplicons. BMC bioinformatics. 2012;13(10):1.
398. Campo DS, Dimitrova Z, Yamasaki L, Skums P, Lau DT, Vaughan G, et al. Next-generation sequencing reveals large connected networks of intra-host HCV variants. BMC genomics. 2014;15(5):1.
399. Batagelj V, Mrvar A. Pajek—analysis and visualization of large networks: Springer; 2004.
400. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution. 2013:mst197.
401. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-91.
402. Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Computer applications in the biosciences: CABIOS. 1993;9(6):745-56.
403. Bukh J, Miller RH, Purcell RH, editors. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Seminars in liver disease; 1995.
404. Keohavong P, Thilly WG. Fidelity of DNA polymerases in DNA amplification. Proceedings of the National Academy of Sciences. 1989;86(23):9253-7.
405. Mullan B, Sheehy P, Shanahan F, Fanning L. Do Taq-generated RT-PCR products from RNA viruses accurately reflect viral genetic heterogeneity? Journal of viral hepatitis. 2004;11(2):108-14.
406. Ribeiro RM, Li H, Wang S, Stoddard MB, Learn GH, Korber BT, et al. Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate. PLoS Pathog. 2012;8(8):e1002881.
407. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC evolutionary biology. 2011;11(1):131.
408. Belshaw R, Gardner A, Rambaut A, Pybus OG. Pacing a small cage: mutation and RNA viruses. Trends in Ecology & Evolution. 2008;23(4):188-93.
409. Schmidt-Martin D, Crosbie O, Kenny-Walsh E, Fanning LJ. Hepatitis C quasispecies adaptation in the setting of a variable fidelity polymerase. 2012.
410. Weiner AJ, Geysen HM, Christopherson C, Hall JE, Mason TJ, Saracco G, et al. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. Proceedings of the National Academy of Sciences. 1992;89(8):3468-72.
411. Thomssen R, Bonk S, Thiele A. Density heterogeneities of hepatitis C virus in human sera due to the binding of β-lipoproteins and immunoglobulins. Medical microbiology and immunology. 1993;182(6):329-34.
412. Smith DB, Simmonds P. Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. Journal of molecular evolution. 1997;45(3):238-46.
413. Lok AS, McMahon BJ. AASLD practice guideline update. Hepatology. 2009.

414. Aronsohn A, Jensen D. Informed deferral: a moral requirement for entry into the hepatitis C virus treatment warehouse. Hepatology. 2012;56(5):1591-2.
415. Beinhardt S, Aberle JH, Strasser M, Dulic–Lakovic E, Maieron A, Kreil A, et al. Serum level of IP-10 increases predictive value of IL28B polymorphisms for spontaneous clearance of acute HCV infection. Gastroenterology. 2012;142(1):78-85. e2.
416. Darling JM, Aerssens J, Fanning G, McHutchison JG, Goldstein DB, Thompson AJ, et al. Quantitation of pretreatment serum interferon-γ–inducible protein-10 improves the predictive value of an IL28B gene polymorphism for hepatitis C treatment response. Hepatology. 2011;53(1):14-22.
417. Fattovich G, Covolo L, Bibert S, Askarieh G, Lagging M, Clément S, et al. IL28B polymorphisms, IP-10 and viral load predict virological response to therapy in chronic hepatitis C. Alimentary Pharmacology & Therapeutics. 2011;33(10):1162-72.
418. Kumthip K, Pantip C, Chusri P, Thongsawat S, O'Brien A, Nelson K, et al. Correlation between mutations in the core and NS5A genes of hepatitis C virus genotypes 1a, 1b, 3a, 3b, 6f and the response to pegylated interferon and ribavirin combination therapy. Journal of viral hepatitis. 2011;18(4):e117-e25.
419. Okanoue T, Itoh Y, Hashimoto H, Yasui K, Minami M, Takehara T, et al. Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study. Journal of gastroenterology. 2009;44(9):952-63.
420. Hill A, Cooke G. Hepatitis C can be cured globally, but at what cost. Science. 2014;345(6193):141-2.