UNIVERSITY OF CYPRUS

SCHOOL OF PURE AND APPLIED SCIENCES
DEPARTMENT OF BIOLOGICAL SCIENCES

LABORATORY OF BIOTECHNOLOGY AND MOLECULAR VIROLOGY

(ASSOCIATE PROFESSOR LEONDIOS G. KOSTRIKIS)

GENETIC STUDIES ON THE HEPATITIS C VIRUS

Ph.D. **DISSERTATION**

VICTORIA L. DEMETRIOU

NICOSIA 2010

Details of Doctoral Programme

1. **Date of Entrance to Ph. D. Programme:** 4th September 2006

2. **Research Advisor:** Leondios G. Kostrikis, Associate Professor

3. **Date of Qualifying Exam:** 1st December 2008

4. **Date of Ph. D. Thesis Proposal Presentation:** 13th August 2009

5. **Members of Advisory Committee:**

   I. Leondios G. Kostrikis, Associate Professor, Department of Biological Sciences, University of Cyprus
   II. Vassileios J. Promponas, Lecturer, Department of Biological Sciences, University of Cyprus (presiding)
   III. Katerina Strati, Lecturer, Department of Biological Sciences, University of Cyprus
   IV. Anna Papa-Konidari, Associate Professor, Department of Microbiology, Aristotle University of Thessalonica, Greece
   V. Peter Karayiannis, Reader in Molecular Virology, Faculty of Medicine, Imperial College London, UK

6. **Date of Ph. D. Thesis Defence:** 6th December 2010

Signature of Research Advisor

To my parents,

Demetris and Susan Demetriou

Abstract

Hepatitis C virus (HCV) infection is a major cause of chronic liver disease, cirrhosis and death. It is estimated that 180 million people are infected with HCV worldwide. It is a highly genetically diverse virus, separated into 6 confirmed genotypes and multiple subtypes. Sequence variation is caused mainly by high rate of replication and the lack of proof reading activity in the viral RNA polymerase, and, less frequently, by recombination events. Determination of genotype and knowledge of the sequence diversity are important for studies on molecular epidemiology and evolution, for classification and nomenclature, and for clinical management.

This thesis is concerned with the investigation of the genetic variability of HCV, and is separated into two sections. In the first part, the molecular epidemiology of HCV infection in Cyprus is presented for the first time among the general population and within high-risk groups. A protocol for the amplification and sequencing of two genomic regions in all genotypes was designed. Phylogenetic analysis of the resulting sequences revealed high genetic heterogeneity, a polyphyletic infection, multiple points of introduction, and the existence of recombinant and unclassified strains on the island. All HCV genotypes were discovered, except genotype 6. The most prevalent genotype was 1, followed by 3, 4, 2 and 5. The high risk groups only exhibited strains belonging to 3a, 1b and 1a. Certain monophyletic clusters among their sequences demonstrate needle or syringe sharing between individuals, but the infection in injecting drug users on the island seems to be running in parallel to that of the general population.

The second part of the thesis describes the characterisation of uncommon HCV variants found in the general population. This was done using strain-specific experimental design for amplification and sequencing of multiple PCR products spanning the near-full genome. Sequence comparisons and phylogenetic analyses gave the following results: Two 2k/1b strains were identified, significantly increasing the 2k/1b sequence data in the databases, confirming the spread of this strain among intravenous drug users and its association with countries of the former Soviet Union, and implying a more complicated evolutionary history that that currently held. One unclassified strain appearing to belong to genotype 1 was discovered here for the first time and a near-full reference sequence is provided. The strain is genetically distant from other subtypes and is on the verge of belonging to a new genotype.

Two strains of a previously discovered unclassified genotype 4 type were sequenced along the near-full genome for the first time, providing reference sequences for this variant, possibly novel subtype 4v. Considering the dominance of 1a and 1b strains in public databases, these strains will all provide valuable sequence information for research purposes.

The molecular epidemiology of HCV in Cyprus and near-full genome sequences of unclassified novel strains are presented here for the first time, highlighting the current drift of certain subtypes into Europe, and the significance of intravenous drug use and immigration. Better knowledge of the global HCV sequence diversity is needed to understand the evolutionary diversity of the epidemic and guide efforts for vaccine and drug development, and near-full genome reference sequences of uncommon subtypes are required for better HCV classification, molecular evolutionary studies and optimised treatment.

Περίληψη

Η μόλυνση με τον ιό της ηπατίτιδας C (HCV) είναι ένας από τους κύριους παράγοντες για χρόνια ηπατική πάθηση, κίρρωση και θάνατο. Υπολογίζεται πως 180 εκατομμύρια άνθρωποι παγκοσμίως είναι μολυσμένοι με το τον ιό αυτό. Είναι ένας γενετικά ποικιλόμορφος ιός, που χωρίζεται σε 6 γονότυπους και πολλούς υπότυπους. Η ποικιλομορφία στο γονιδίωμα του προκαλείται κυρίως από τον υψηλό ρυθμό αναπαραγωγής και την έλλειψη της ιδιότητας να διορθώνει λάθη στην RNA πολυμεράση, και λιγότερο συχνά, από ανασυνδυασμούς. Ο προσδιορισμός του γονότυπου και οι γνώσεις για την ετερογένεια ανάμεσα σε αλληλουχίες είναι σημαντικά για μελέτες μοριακής επιδημιολογίας και εξέλιξης, για την ταξινόμηση και ονομασία, και για την κλινική αντιμετώπιση.

Το κείμενο αυτό ασχολείται με την μελέτη της γενετικής ποικιλότητας του HCV και χωρίζεται σε δύο μέρη. Στο πρώτο μέρος περιγράφεται για πρώτη φορά η μοριακή επιδημιολογία της HCV λοίμωξης στην Κύπρο ανάμεσα στο γενικό πληθυσμό και σε ομάδες υψηλού κινδύνου. Σχεδιάστηκε πρωτόκολλο για πολλαπλασιασμό και αλληλούχιση δύο γονιδιωματικών περιοχών, που καλύπτει όλους τους γονότυπους. Η φυλογενετική ανάλυση που έγινε με τις αλληλουχίες αυτές αποκάλυψε ψηλή γενετική ετερογένεια, πολυφυλετική λοίμωξη, πολλαπλά σημεία εισδοχής και την ύπαρξη ανασυνδυασμένων μορφών και μορφών που δεν ταξινομούνται με τους γνωστούς υπότυπους. Βρέθηκαν στελέχη από όλους τους γονότυπους, εκτός από τον 6. Ο κυρίαρχος γονότυπος ήταν ο 1, και μετά οι 3, 4, 2 και 5. Οι ομάδες υψηλού κινδύνου παρουσίασαν στελέχη μόνο από τους υπότυπους 3a, 1b και 1a. Η ύπαρξη κάποιων μονοφυλετικών σχέσεων με στελέχη από αυτές τις ομάδες δείχνει τη χρήση χρησιμοποιημένων συριγγών ή βελόνων ανάμεσα σε κάποια άτομα, αλλά η λοίμωξη στους χρήστες ενδοφλέβιων ναρκωτικών φαίνεται να κινείται παράλληλα με αυτή στο γενικό πληθυσμό.

Το δεύτερο μέρος του κειμένου περιγράφει τον χαρακτηρισμό σπάνιων στελεχών HCV που βρέθηκαν στο γενικό πληθυσμό. Αυτό έγινε με πρωτόκολλο, το οποίο σχεδιάστηκε για το κάθε στέλεχος ξεχωριστά, για πολλαπλασιασμό και αλληλούχιση αλληλεπικαλυπτόμενων προϊόντων PCR, που να εκτείνονται σε σχεδόν ολόκληρο το γονιδίωμα. Η σύγκριση αλληλουχιών και η φυλογενετική ανάλυση είχαν ως αποτέλεσμα τα ακόλουθα: Δύο στελέχη 2k/1b εξακριβώθηκαν, αυξάνοντας σημαντικά τα δεδομένα αλληλουχιών για το 2k/1b στις διεθνείς βάσεις δεδομένων, επιβεβαιώνοντας την εξάπλωση

αυτού του είδους ανάμεσα σε χρήστες ενδοφλέβιων ναρκωτικών και την συσχέτισή του με χώρες της πρώην Σοβιετικής Ένωσης, και υποδηλώνοντας μια εξελικτική ιστορία πιο πολύπλοκη από ότι έχει τεθεί μέχρι σήμερα. Ένα στέλεχος που φαίνεται να ανήκει στον γονότυπο 1, αλλά που δεν ταξινομείται με κανένα γνωστό υπότυπο, ανακαλύφθηκε εδώ για πρώτη φορά και κατατίθεται η αλληλουχία του σχεδόν ολόκληρου γονιδιώματος. Το στέλεχος είναι γενετικά απομακρυσμένο από τους άλλους υπότυπους και θα μπορούσε να αντιπροσωπεύει και καινούριο γονότυπο. Δύο στελέχη ενός είδους που ανήκει στον γονότυπο 4 αλλά δεν ταξινομείται σε κάποιο γνωστό υπότυπο, αλληλουχήθηκαν σε σχεδόν ολόκληρο το γονιδίωμα για πρώτη φορά, παρέχοντας στελέχη αναφοράς για αυτό το είδος, που πιθανά ανήκει σε νέο υπότυπο, τον 4v. Έχοντας υπόψη την κυριαρχία των υποτύπων 1a και 1b σε βάσεις δεδομένων, τα στελέχη αυτά θα παρέχουν πολύτιμες πληροφορίες για ερευνητικούς σκοπούς.

Η μοριακή επιδημιολογία της HCV λοίμωξης στην Κύπρο και η αλληλούχιση στο σχεδόν ολόκληρο γονιδίωμα καινούριων στελεχών παρουσιάζονται εδώ για πρώτη φορά, υπογραμμίζοντας την πρόσφατη μετατόπιση κάποιων υποτύπων προς την Ευρώπη και τονίζοντας το σημαντικό ρόλο των χρηστών ενδοφλέβιων ναρκωτικών και των μεταναστών. Χρειάζεται πιο λεπτομερής γνώση για την ετερογένεια του HCV ανά το παγκόσμιο για να γίνει κατανοητή η εξελικτική ποικιλομορφία της επιδημίας και να καθοδηγήσει τις προσπάθειες για ανακάλυψη εμβολίου και ανάπτυξη φαρμάκων. Στελέχη αναφοράς που να είναι γνωστή ολόκληρη η αλληλουχία τους για πιο σπάνια είδη είναι σημαντικά για καλύτερη ταξινόμηση, μοριακές μελέτες, και βελτιστοποίηση της θεραπείας.

Acknowledgements

I would like to express my gratitude to all those who helped in the completion of this thesis. First and foremost, I express my thanks to my supervisor Leondios Kostrikis for giving me the opportunity to carry out research and complete a Ph. D. project in his laboratory, and for his supervision and guidance throughout.

I am obliged to my many collaborators who have taken part in various sections of this work: David van de Vijver, who is a pleasure to work with, and has provided his time, assistance and supportive comments on my work and construction of manuscripts; Vassilis Promponas, Peter Karayiannis, Anna Papa-Konidari, Katerina Strati and Georgios Pollakis, for their guidance, constructive criticism and helpful comments in my qualifying exam, thesis proposal and thesis defence; the clinicians and collaborators who have participated in a pancyprian network of HCV sample collection and have shown enthusiasm and support for our research in this area: Solonas Cheimonides, Andreas Evgeniou, Emmeleia Hadjigeorgiou-Vounou, Maria Koliou-Mazeri, Panagiotis Papakyriakou, Loukas Petsas, Georgios Potamitis, Natasa Savvopoulou and Neoklis Georgiades at EKTEPN, the managers of the rehabilitation centres for drug abusers and Dr. Katerina Hadjivasileiou at the State Correctional Facilities. Their help and collaboration made possible an investigation in very difficult cohorts. Also, I thank the staff at the sampling centres and C. Charilaou Biomedical Laboratory for their help in sample collection and testing. I am especially indebted to all the participants of the study for their kind cooperation during a difficult time of their life. This study could not have been completed without them.

I would like to give special thanks to my friends and colleagues from the Laboratory of Biotechnology and Molecular Virology Ioanna Kousiappa, Johana Hezka, Erika Loizidou, Andreas Hadjinicolaou, Katie Charalambous, Eftychia Kyriakou, Sarah Gilliland and Efthymia Yiacoumi for their moral, technical and intellectual support and their friendship during my time at the lab.

Furthermore, I acknowledge the Cyprus Research Promotion Foundation (grant number ΕΝΙΣΧ/0506/34), Birch Biomedical, the Cyprus Anti-drugs Council (item 633 of the Budget for Research Programmes 2008) and the University of Cyprus for providing financial support for myself and the work carried out for my dissertation.

Finally, I am deeply indebted to my close friends and family for their unwavering support, encouragement and loving patience throughout these past four and half years.

Table of Contents

List of Figures and Tables

Figures

Tables

1. Introduction

1.1. Hepatitis C infection

Hepatitis C virus (HCV) infection is a life-shortening liver disease associated with complex and expensive morbidity and decreased quality of life, a major contributor to liver cirrhosis and hepatocellular carcinoma (HCC) worldwide, and the most common indication for liver transplantation in developed countries (Alter, 1997; Charlton, 2001; Wasley and Alter, 2000). It is recognised as a major global public health problem, infecting an estimated 180 million people, comprising ~3% of the global population, and although the incidence of hepatitis C infection has dramatically decreased during the past decade, millions more are still newly infected every year (Sy and Jamal, 2006). Accurate numbers, however, are lacking, in particular for developing countries where most of the world's population resides, and so global estimates of the prevalence are difficult to define (Rinaldi, 2006).

The prevalence of HCV varies widely with geographical location and within populations. The overall prevalence of HCV is almost 1% in the United Kingdom, 2% in North America, 3-4% in some Mediterranean and Asian countries and more than 10% in parts of Central Africa and Egypt (Health Protection Agency, 2008; Wasley and Alter, 2000). In fact, over 85% of the world's HCV-infected subjects exist in regions of Africa, Southeast Asia and Middle Eastern countries. In particular, HCV is highly prevalent in Egypt with more than 19% of the population infected and chronic HCV representing one of the top five leading causes of death, due to the unintended transmission of HCV during a national campaign for the parenteral treatment of schistosomiasis (Frank *et al.*, 2000). The current state of HCV infection predicts a dramatic global rise of HCV-related cirrhosis and death from HCC in the next three decades (Deuffic-Burban *et al.*, 2006; Deuffic-Burban *et al.*, 2007; Sweeting *et al.*, 2007). Therefore, HCV is, and will continue to be, a major drain on healthcare resources worldwide, highlighting the urgent need for intervention to minimise the burden of disease in those already infected and to reduce the incidence of new infections (Thomson, 2009).

Hepatitis C has been termed the "silent pandemic" because it may cause no noticeable symptoms for decades as it slowly destroys the liver. Hence, in most cases during this asymptomatic period, the carrier does not know they are infected and will not present for medical care until it is too late for effective therapy and during which time they are likely to have unknowingly transmitted the infection to others, who will in turn act in the same way.

Consequently, early hepatitis C infection is infrequently diagnosed. Acute infection can be defined as HCV RNA detection in serum in the absence of detectable antibodies. HCV RNA can be detected in the serum within 1 to 2 weeks after exposure (Thimme *et al.*, 2002). Most patients acutely infected with HCV become antibody positive within 3 months, although seroconversion can sometimes take 6 months or longer (Farci *et al.*, 1991). In about 15-20% of acute HCV infections, the patient will recover spontaneously, but in most cases the disease will progress to chronicity, and the patients are at risk for the development of extrahepatic manifestations, compensated and decompensated cirrhosis, and HCC. Chronic hepatitis C is marked by the persistence of HCV RNA in the blood for at least 6 months after onset of acute infection.

The chronic nature of hepatitis C infection influences the clinical approach and management of this disease. Cirrhosis develops in an estimated 10% to 15% of individuals with chronic HCV infection, and this usually happens within the first 20 years. The rate of progression to cirrhosis is highly variable, and is influenced by several external and host factors, such as the amount of alcohol consumption, older age at infection, male gender, the degree of inflammation and fibrosis present on liver biopsy, jaundice, coinfection with human immunodeficiency virus (HIV) or hepatitis B virus (HBV), and comorbid conditions (reviewed by Chen and Morgan, 2006; Seeff, 2002; Thomson, 2009). The liver biopsy is the gold standard for the grading and staging of chronic hepatitis C.

An understanding of the natural history and long-term outcomes of HCV infection is essential for clinicians in order to identify patients at risk of HCV-related complications and offer effective management, treatment, and counselling.

1.2. Transmission

Hepatitis C is a blood-borne infection, and until the introduction of diagnostic screening in 1991, the virus was transmitted mainly through blood, blood products, haemodialysis, and organ transplantation (Alter, 2002; Alter, 2007; Memon and Memon, 2002). Once blood donation screening by both antibody and nucleic acid testing became standard practice in developed countries, transfusion-related HCV transmission was significantly reduced. However, in developing countries unsafe medical practices, therapeutic injections, inadequate disinfection practices, non-sterile medical and dental procedures, and unscreened blood transfusions may still account for significant HCV transmission and serve as

a bridge to the general population. The transmission is also associated with additional percutaneous exposures such as unsanitary skin scarification, piercing and tattooing.

HCV infection in the Western world now occurs predominantly by intravenous drug use through sharing of needles or other injecting equipment, and is considered the current major risk factor and reservoir for infection (Aceijas and Rhodes, 2007; Aitken *et al.*, 2008; Crofts *et al.*, 1999; Esteban *et al.*, 2008; Judd *et al.*, 2005; Micallef *et al.*, 2007; Pybus *et al.*, 2005; Shepard *et al.*, 2005; Sweeting *et al.*, 2009; Sy and Jamal, 2006). Injecting drug users constitute by far the largest pool of HCV infection in industrialised countries. Prevalence varies between 27% and 74% in different cohorts, which is extremely high (Hagan *et al.*, 2007; Sutton *et al.*, 2008; Thomson, 2009). Users who have practising injecting drugs for many years display the highest prevalence of HCV infection, whereas short- term users exhibit prevalence rates of 20-46%, which are still significantly high (Shepard *et al.*, 2005; Sweeting *et al.*, 2009). Despite these very high rates of infection, antiviral therapy has not been consistently offered to this population, as they are at high risk of re-infection and are not considered to represent a cost-effective use of high cost combination therapy. Strategies for the reduction of the burden of HCV infection in this extremely high-risk population are therefore essential, as they are the population in which the majority of incident infections occur. Due to the high prevalence of such high-risk behaviour and others, such as unsanitary tattoo and piercing practises, and getting into fights frequently, imprisoned populations are also considered at high risk for HCV infection and transmission (Holsen *et al.*, 1993; McGovern *et al.*, 2006; Miller *et al.*, 2009a; Reindollar, 1999; Rhodes *et al.*, 2008; Spaulding *et al.*, 2006; Weinbaum *et al.*, 2005; Zalumas and Rose, 2003).

Unlike HIV and HBV, sexual transmission of HCV is very infrequent (Vandelli *et al.*, 2004). Such methods of spread of HCV infection have been reported recently (Ghosn *et al.*, 2008; Richardson *et al.*, 2008; Urbanus *et al.*, 2009; van de Laar *et al.*, 2007), but sexual behaviour and HCV transmission has been a controversial issue for many years. In the USA and Europe there appears to be an epidemic of acute HCV emerging among HIV-infected men who have sex with men (MSM), as opposed to heterosexual partners, as transmission in this group is more likely to be permucosal rather than percutaneous, and correlates with the numbers of sexual partners, the sharing of drugs through the nasal or anal route and high-risk sexual practices (Danta *et al.*, 2007).

Vertical transmission of HCV has been recorded but is much less frequent than other routes of transmission, but the likelihood appears to double in women coinfected with HIV

(Chang, 1996; Dowd *et al.*, 2008; England *et al.*, 2006; Ferrero *et al.*, 2003; Indolfi and Resti, 2009; Ngo-Giang-Huong *et al.*, 2010; Parthiban *et al.*, 2009; Pembrey *et al.*, 2005; Rapicetta *et al.*, 2000; Shebl *et al.*, 2009; Weiner *et al.*, 1993).

As discussed earlier, newly HCV-infected individuals are often asymptomatic for many years before presenting for medical care and realising they are carriers of the virus. This fact presents the main danger for transmission of this disease, giving it a very large window of opportunity to spread to multiple other hosts.

1.3. HCV genome organisation and life cycle

It has now been over 20 years since the identification of HCV, the aetiological agent of hepatitis C (Choo *et al.*, 1989). Historical accounts of this discovery have been described in various review papers (Alter, 1999; Farci, 2002; Houghton, 2009). Briefly, in the 1970´s many cases of parenterally-transmitted hepatitis were found to be due to neither hepatitis A nor B (Feinstone *et al.*, 1975). The unknown infectious agent was therefore then named non-A, non-B hepatitis (NANBH). Many laboratories throughout the world made extensive efforts to characterise the NANBH aetiological agent, but this proved extremely difficult (Shih *et al.*, 1986). The causative agent of NANBH was finally identified in 1989, following the isolation of a small clone using a blind cDNA immunoscreening approach (Choo *et al.*, 1989) and henceforth named hepatitis C virus.

HCV is a small, enveloped virus belonging to the *Hepacivirus* genus within the family *Flaviviridae*, consisting of a single-stranded, positive-sense RNA genome approximately 9,600 nucleotides long. Following the identification of the first complete HCV genome, its genetic organisation could then be elucidated (Choo *et al.*, 1991; Kato *et al.*, 1990). The viral genome is composed of a single open reading frame (ORF) encoding a polyprotein precursor of about 3000 amino acids, flanked by 5'- and 3'-non-coding regions (NCR), of 341 and approximately 230 nucleotides in length, respectively (Figure 1). Both NCRs bear highly conserved RNA structures essential for polyprotein translation and genome replication. The HCV polyprotein is co- and post-translationally processed by cellular and viral proteases at the level of the endoplasmic reticulum (ER) to yield ten mature structural and non-structural (NS) proteins. HCV structural proteins include the core nucleocapsid protein and the envelope glycoproteins E1 and E2. They are separated from the non-structural proteins by the short membrane polypeptide p7. The NS proteins include the NS2 protease, NS3 serine protease/RNA helicase, the NS4A protease cofactor polypeptide, NS4B, NS5A

phosphoprotein and NS5B RNA-dependent RNA polymerase (RdRp) (Dubuisson, 2007; Penin *et al.*, 2004). The existence of one or more previously unknown HCV proteins potentially produced due to ribosomal frameshifts has been suggested recently (Choi *et al.*, 2003; Varaklioti *et al.*, 2002; Walewski *et al.*, 2001; Xu *et al.*, 2001). The NS proteins and p7, reported to have ion channel activity, are essential for polyprotein processing and viral replication and have become the centre of attention for intensive drug development and clinical testing work for the last 1-2 decades (de Bruijne *et al.*, 2009a; De Francesco *et al.*, 2003; Dubuisson, 2007; Harrison, 2007; Manns *et al.*, 2007; Meier and Ramadori, 2009; Moradpour *et al.*, 2002; Parfieniuk *et al.*, 2007; Pawlotsky, 2005; Sherman *et al.*, 2007; Webster *et al.*, 2009). Tremendous progress has been made in the characterisation of HCV component structures, despite the lack of a productive cell culture system until recently. These insights have helped in understanding the mechanisms of HCV replication and pathogenesis. Most importantly, detailed knowledge of the structure of key viral proteins holds the promise

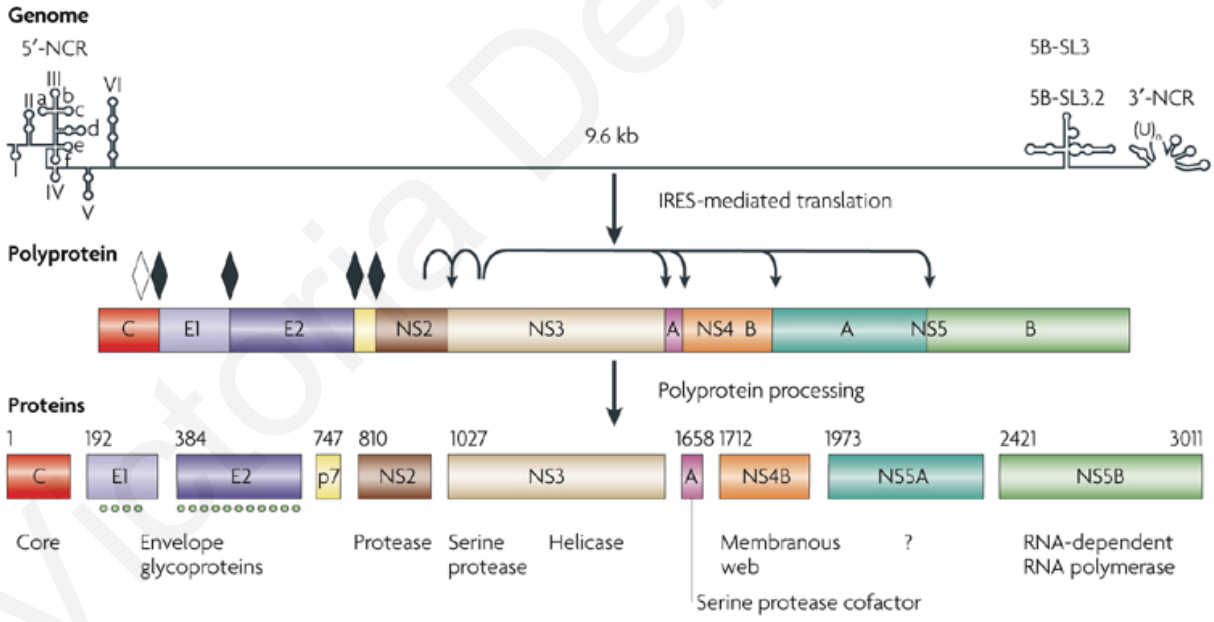

Figure 1. A schematic depiction of the HCV genome is seen at the top, with simplified RNA secondary structures in the 5'- and 3'-NCRs, the core region and the NS5B stem-loop 3 *cis*-acting replication element (5B-SL3). In the middle, the polyprotein precursor is represented, and this is processed into the mature structural and non-structural viral proteins (bottom). Numbers above each protein denote amino-acid locations, according to strain H77 (GenBank Acc. No. AF009606). The diamonds indicate cleavage sites of the polyprotein by host peptidases, and the arrows point out sites cleaved by the HCV NS2/3 and NS3/4A proteases. Dots in E1 and E2 indicate the glycosylation of the envelope proteins. Reproduced from Moradpour *et al.*, 2007.

that efficient and specific inhibitors of HCV replication could be developed within the foreseeable future.

HCV virions circulate in various forms in the blood of the infected host, but the infectious populations appear to be the forms bound to very-low-density lipoproteins (VLDL) and low-density lipoproteins (LDL) (Monazahian *et al.*, 2000). Electron microscopy studies of HCV particles have been hindered by the low amount of virus particles available in serum and the lack of an efficient cell culture system. HCV is thought to adopt a classical icosahedral scaffold with a 50-60 nm diameter (Wakita *et al.*, 2005) in which its E1 and E2 glycoproteins are anchored to a lipid envelope, derived from the host cell. Underneath the membrane is the nucleocapsid that likely is composed of multiple copies of the core protein, forming an internal icosahedral viral coat encapsidating the viral genomic RNA. E1 and E2 are essential for host-cell entry, by binding to receptors and inducing fusion between the viral envelope and the host cell membrane (Bartosch *et al.*, 2003).

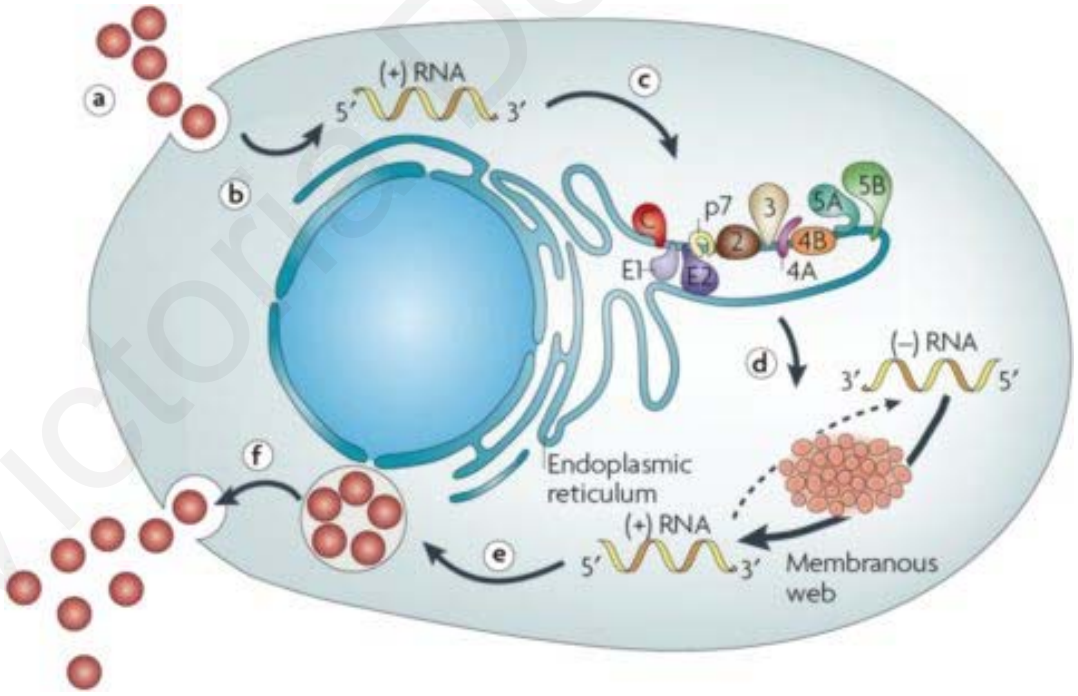

Figure 2. HCV life cycle. (a) virus attachment and internalisation, (b) cytoplasmic release and uncoating, (c) IRES-mediated translation and polyprotein processing, (d) RNA replication, (e) packaging and assembly, (f) virion maturation and release. The topology of the HCV proteins at the ER membrane is shown schematically. Reproduced from Moradpour *et al.*, 2007.

Two decades after the discovery of this virus, many central questions concerning the HCV life cycle and host interactions remain unsolved (Chevaliez and Pawlotsky, 2006). A simplified diagram of the current knowledge on this is seen in Figure 2. The main target of HCV is the hepatocyte, in which it replicates. Entry of the HCV virion into the host cell is a complex and highly regulated process, not yet fully understood. Initial attachment of the virion to the cell may require the presence of both glycosaminoglycans and the LDL receptor (Agnello *et al.*, 1999; Barth *et al.*, 2003; Monazahian *et al.*, 1999). Entry then proceeds through interaction with the scavenger receptor class B type I (SR-BI), the tetraspanin CD81, occludin, and the tight junction proteins Claudin-1, -6 or -9 (Dubuisson *et al.*, 2008; Evans *et al.*, 2007; McKeating *et al.*, 2004; Pileri *et al.*, 1998; Ploss *et al.*, 2009; Scarselli *et al.*, 2002; Zeisel *et al.*, 2007). Most of these potential HCV receptors have been shown to be necessary, but not sufficient, for viral entry. It is important to remember, however, that the process of viral binding and entry may vary with the details of the experimental model used and with the environment of the virus-host cell interaction *in vivo*. The dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN or CD209) and the liver/lymph node-specific intercellular adhesion molecule-3 (ICAM-3)-grabbing integrin (L-SIGN or CD209L) have also been proposed as tissue-specific capture receptors for HCV present in various cell types that could play a critical role in viral pathogenesis and tissue tropism (Gardner *et al.*, 2003; Lozach *et al.*, 2004; Lozach *et al.*, 2003; Pöhlmann *et al.*, 2003).

After attachment to the cell surface, fusion between the viral and cellular membranes leads to the release of the HCV RNA genome into the cytosol through removal of the nucleocapsid proteins. This is controlled by the viral surface glycoproteins, believed to belong to class II fusion proteins (Yagnik *et al.*, 2000). The 5'NCR of the viral RNA contains an internal ribosome entry site (IRES) that binds the 40S ribosomal subunit and initiates polyprotein translation in a cap-independent manner. The resulting polyprotein is targeted to the ER membrane and cleaved by viral and host proteases to produce the mature viral proteins (Bartenschlager and Lohmann, 2000; Lindenbach and Rice, 2005; Penin *et al.*, 2004). Expression of the HCV proteins leads to rearrangements of intracellular membranes to form a replication complex that associates all the viral NS proteins, cellular components and nascent RNA strands, creating a structure called the membranous web (Bartenschlager *et al.*, 2004; Egger *et al.*, 2002; Kapadia and Chisari, 2005) The elucidation of this process was possible due to the recent development of an infectious cell culture replicon system (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005; Zhong *et al.*, 2005). This strategy may help in compartmentalisation

and local concentration of viral products, physical support and organisation of the RNA replication complex, tethering of viral RNA during unwinding, supply of lipid constituents important for replication, and protection of viral RNA from double-strand RNA-mediated host defences and RNA interference. HCV replication occurs in two steps catalysed by the NS5B RdRp. During the first step, the positive-strand genome RNA serves as a template for the synthesis of a negative-strand intermediate. In the second step, negative-strand RNA serves as a template to produce numerous positive-strand RNA molecules that will subsequently be used for polyprotein translation, synthesis of new intermediates of replication or packaging into new virus particles (Bartenschlager *et al.*, 2004). The positive-strand RNA progeny is transcribed in a five to ten fold excess compared to negative-strand RNA.

Accumulation of HCV genomic RNA and the structural proteins leads to the assembly of a nucleocapsid, which acquires its envelope within an intracellular compartment, and the virion is then secreted via the classical secretory pathway. Little is known about the mechanisms of HCV nucleocapsid assembly and release, due to lack of appropriate study models.

1.4. HCV genetic heterogeneity, nomenclature and classification

The HCV genome is noted for its high genetic variability, and exhibits substantial global sequence diversity. This phenomenon is caused by its rapid replication rate and the lack of proof-reading activity of its RNA polymerase. In a single patient with chronic HCV infection, the turnover rate is about 10^{12} virions per day, leading to a mutation rate estimated at 10^{-3}-10^{-4} base substitutions per site per year (Bartenschlager and Lohmann, 2000; Bukh *et al.*, 1995; Sentandreu *et al.*, 2008). This high genetic heterogeneity works as an advantage to the virus, giving it its incredible propensity to escape host immune responses and selective pressure by antiviral drugs, enabling it to persist for the lifetime of the host.

The heterogeneity of the HCV genome has lead to a proposed consensus of six genotypes (1-6) and numerous closely related subtypes (a, b, c,...) based on sequence variation in the 5'NCR, core, E1 and NS5B regions (Robertson *et al.*, 1998; Sentandreu *et al.*, 2008; Simmonds, 1999; Simmonds *et al.*, 2005; Simmonds *et al.*, 1993). In addition to genotypes, HCV exists within its hosts as a pool of constantly evolving, genetically distinct but closely related variants referred to as quasispecies (Martell *et al.*, 1992; Weiner *et al.*, 1991). New variants still continue to be identified, usually for genotype 4 and 6 (Koletzki *et al.*, 2009; Li *et al.*, 2006; Li *et al.*, 2009a; Li *et al.*, 2009b; Noppornpanth *et al.*, 2008; Wang *et al.*, 2009; Xia

et al., 2008a) as a result of long endemicity of these genotypes in certain geographical areas (Pybus *et al.*, 2001). Furthermore, a possible new genotype (7a) has been discovered recently (Murphy *et al.*, 2007a). At the level of complete genomes, HCV exhibits around 30-35% nucleotide variation between genotypes and 20-25% variation in subtypes of the same genotype (Simmonds *et al.*, 2005). Intensive sequencing of HCV genomes is currently being conducted by labs all over the world, and over 90,000 HCV sequences, including 1,200 full-length genomes, have been deposited so far in generic data banks (GenBank, http://www.ncbi.nlm.nih.gov/genbank). However, these figures are significantly biased towards 1a and 1b sequences, because they are the most prevalent in developed countries, where most HCV research is carried out. Databases dedicated specifically to HCV sequences are publicly available, allowing sequence classification, annotation and analysis of large collections. Such databases are the Los Alamos HCV sequence database (http://www.hcv.lanl.gov), the European HCV database (http://euhcvdb.ibcp.fr) and the Japanese HCV database (http://s2as02.genes.nig.ac.jp).

The genotype is presently a major factor in both the choice of treatment and prognosis (Zeuzem, 2004). The relationship between HCV sequence variability and liver disease status or resistance to current antivirals is unclear and is probably multifactorial (Pawlotsky, 2003a; Pawlotsky *et al.*, 1999; Pawlotsky *et al.*, 1998). The realisation of the diversity of HCV has important clinical implications since 80-90% of patients infected with genotypes 2 and 3 are curable with the current standard therapy (a combination of pegylated interferon-α and ribavirin), whilst only about 40-50% of patients infected with genotypes 1 and 4 are cured in this way.

Until 2001, HCV was thought to evolve in a clonal manner, with diversity generated through the accumulation of mutations. It has even been suggested that any recombination events are very rare in vivo and that the resulting virions are not viable (Simmonds *et al.*, 1994; Smith and Simmonds, 1997). However, homologous recombination has been demonstrated between different genotypes or different subtypes of a genotype. A 2k/1b recombinant was found in St. Petersburg, Russia (Kalinina *et al.*, 2002), and seems to be circulating among intravenous drug users in Russia. Isolates of this strain have also been found in Ireland, Estonia, Uzbekistan, Western Siberia and France (Kurbanov *et al.*, 2008a; Kurbanov *et al.*, 2008b; Moreau *et al.*, 2006; Morel *et al.*, 2010; Tallo *et al.*, 2007). Several other events of recombination have also been reported to date from various parts of the world (Colina *et al.*, 2004; Cristina and Colina, 2006; Kageyama *et al.*, 2006; Legrand-Abravanel *et*

al., 2007; Moreno *et al.*, 2006; Moreno *et al.*, 2009; Noppornpanth *et al.*, 2006; Sentandreu *et al.*, 2008), but are isolated events and do not appear to be circulating in infected populations. HCV recombination break points have been located mainly in the non-structural proteins, but an intratypic recombinant with a break point in the structural region has also been identified (Cristina and Colina, 2006; Moreno *et al.*, 2009). As genotyping of HCV is more commonly based on sequence analysis of a single genomic region, the variety and prevalence of recombinant HCV strains is generally thought to be underestimated, since at least two different genomic regions would be required for their detection.

Therefore, recombination events are now starting to be taken into account as contributors to HCV evolution as well as its high mutation rate. Because of the consequences that recombination events have on virus evolution, vaccine development, treatments, infection management and control programmes, it is clearly very important to precisely define the role of recombination in the evolution of HCV and the creation of genetic variation. Research on HCV recombination events is still in its initial stages and the mechanism of recombination generation is still unknown. Genetic transfer from one genotype to another requires the co-existence of the two genotypes in the same cell, which means the patient must be infected with two or more strains. For the recombinant to spread further, it must not only be viable, but also have an advantage over other strains of the virus.

In general, sequence characteristics of a particular subtype can be found throughout the HCV genome. In order to determine the subtype classification of an unknown strain, phylogenetic analysis based on the full-length of the genome would be the most accurate method. However, most strains to date have been classified based on short sequence analysis of the core, E1 and NS5B regions, and the current consensus for identification and subtype assignment states the requirement of analysis comparing at least two coding regions of the genome (Robertson *et al.*, 1998; Sandres-Saune *et al.*, 2003; Simmonds *et al.*, 2005; Simmonds *et al.*, 1994; Stuyver *et al.*, 1994). The 5'NCR is the most conserved region on the HCV genome and is routinely used for HCV RNA detection and genotyping in diagnostic applications, but it does not allow accurate classification of the subtype level (Stuyver *et al.*, 1996; Tamalet *et al.*, 2003). Hence, other, less conserved, and therefore more discriminative genomic regions are required for phylogenetic analysis in order to identify HCV subtypes. For these reasons, the most extensive databases that exist contain partial sequences of the 5'NCR, core, E1 and NS5B regions (http://hcv.lanl.gov).

1.5. Molecular epidemiology and global distribution of HCV

HCV genotype distribution differs by geographic region and by year and mode of transmission (Schroter *et al.*, 2004; Zein, 2000). Although different HCV genotypes appear to have originated in different geographic regions, such as genotypes 1, 2, 4 and 5 in Africa and genotype 3 and 6 in Asia, globalisation is radically changing the face of worldwide HCV epidemiology as a result of modern transmission and human migration (Esteban *et al.*, 2008). Knowledge of genotype distribution in different parts of the world may help clarify the epidemiology and evolution of HCV, and has proven a useful tool for identifying risk groups and distinguishing different routes of transmission (Nakano *et al.*, 2004). Certain modes of transmission are associated with HCV subtype infections, suggesting separate HCV epidemics, but spill-over between different risk groups underlines the value of molecular epidemiological studies to gain insight into the origin and dynamics of HCV infections at the population level.

Geographically, genotypes 1-3 display a worldwide distribution, while genotypes 4-6 have a more restricted geographical pattern and represent long-term endemic infection (Pybus *et al.*, 2001). Genotype 4 is found mainly in North Africa and Mediterranean countries, Egypt in particular, but has recently been spreading to Europe and North America largely via intravenous drug users and immigrants, with a high incidence in Greece (Cenci *et al.*, 2007; Chamberlain *et al.*, 1997a; Chlabicz *et al.*, 2008; Kamal and Nasser, 2008; Katsoulidou *et al.*, 2006; Nguyen and Keeffe, 2005; Nicot *et al.*, 2005; Savvas *et al.*, 2005; Timm *et al.*, 2007; van Asten *et al.*, 2004). Genotype 5 is restricted primarily to South Africa (Chamberlain *et al.*, 1997b; Verbeeck *et al.*, 2006), but has also recently been found in West Flanders, Belgium (Verbeeck *et al.*, 2006), central France (Henquell *et al.*, 2004), and Syria (Antaki *et al.*, 2009), albeit in much smaller numbers. HCV genotype 6 is found in Southeast Asia (Huy and Abe, 2004).

In the geographical area close to Cyprus, in the eastern Mediterranean region, HCV genotypes are not distributed uniformly. In western Turkey, genotype 1b is the most prevalent (Altuglu *et al.*, 2008). Also, a study done in northern Cyprus with civilians, Turkish soldiers, and Northern Cyprus soldiers revealed genotype 1b as the most prevalent (92.4%) (Altindis *et al.*, 2006). In Egypt, the incidence of HCV infection is significantly higher than other countries worldwide and most cases are infected with subtype 4a (Abdel-Hamid *et al.*, 2007). Studies carried out with patients in the Middle East revealed a predominance of HCV

genotypes 4 and 1 (Ramia and Eid-Fares, 2006; Watson *et al.*, 1999). In Greece, genotype 1 is the most prevalent, followed by genotype 3, 4, 2, and 5 (Katsoulidou *et al.*, 2006).

Subtypes 1b and genotype 2 are associated mainly with contaminated blood transfusions and other types of nosocomial transmission, especially in older patients. Among injecting drug users, however, HCV subtypes 1a and 3a predominate, with both subtypes showing an exponential population growth during the 20th century (Cochrane *et al.*, 2002; Morice *et al.*, 2006; Pybus *et al.*, 2005; van Asten *et al.*, 2004). Genotype 3a in particular has been associated significantly with transmission through intravenous drug use in industrialised countries (Bourliere *et al.*, 2002; McCaw *et al.*, 1997; Pawlotsky *et al.*, 1995). It is prevalent mainly in North and South America, Europe, and Australia where practicing intravenous drug abuse is common, and seems to have been transmitted from a common origin through a unique worldwide epidemic that spread rapidly among drug users (Morice *et al.*, 2006; Pybus *et al.*, 2005). As pointed out above, genotype 4 subtypes are also becoming increasingly prevalent in populations of intravenous drug users recently, especially in southern Europe, and its introduction into the European intravenous drug user population seems to be more recent than that of 1a and 3a (Chlabicz *et al.*, 2008; Kamal and Nasser, 2008; van Asten *et al.*, 2004). Finally, one type of recombinant strain, 2k/1b, is also strongly associated with intravenous drug use, and has been found in Russia, Estonia, Uzbekistan, Ireland, and France in individuals from countries of the former Soviet Union (Kalinina *et al.*, 2002; Kurbanov *et al.*, 2008a; Kurbanov *et al.*, 2008b; Moreau *et al.*, 2006; Morel *et al.*, 2010; Tallo *et al.*, 2007).

1.6. HIV/HCV coinfection

Due to sharing common transmission routes, it is very frequent that individuals are simultaneously infected with both HCV and HIV (Kim and Chung, 2009). This type of coinfection is mostly common in haemophiliacs, intravenous drug users, and more recently, MSM (Kretzschmar and Wiessing, 2008; Liu *et al.*, 2008; Miller *et al.*, 2009b; Re *et al.*, 2008; Richardson *et al.*, 2008; Urbanus *et al.*, 2009; van de Laar *et al.*, 2009). The effects of the hepatitis C disease are worse and its natural progression is accelerated in the coinfected population than in HCV monoinfected individuals and conversely, HCV has been associated with a faster progression of HIV to acquired immunodeficiency syndrome (AIDS) (Di Martino *et al.*, 2001; Ragni and Belle, 2001). The consequences of coinfection include increased viral persistence, higher viral loads, a faster rate of progression to liver fibrosis, higher rates of hepatic decompensation, and worse treatment outcomes following interferon-based therapy

(Kim and Chung, 2009; Sulkowski, 2008). There is also the suggestion that some HIV protease inhibitors may be hepatotoxic (Sulkowski, 2008). Additionally, liver transplantation is complicated for this population due to the patients' immunocompromised state (Koziel and Peters, 2007). Despite the increasing prevalence of simultaneous infection, and the increased severity of the combined clinical symptoms, the pathogenesis of HIV/HCV coinfection is still not fully understood.

1.7. Current therapy and new therapeutic targets

The current standard regimen for treating HCV infection includes a 24- or 48-week course of a combination of pegylated interferon-α and ribavirin, depending on the infecting genotype (Fried *et al.*, 2002). A sustained virological response is considered to be the absence of HCV RNA in the peripheral blood of the patient 24 weeks after the treatment has ended. However, this response only occurs in about 80% of patients infected with genotypes 2 and 3, and less than 50% of patients infected with genotypes 1 and 4 (Pawlotsky, 2004; Taylor, 2000; Thomson, 2009). Virological response is therefore dependent on genotype, which may serve as a predictor for treatment duration and response (Ferenci, 2004; Pawlotsky, 2000; Zein, 2000; Zein *et al.*, 1996; Zeuzem, 2004). Not only are there response issues dependant on the HCV genotype, but it has also been reported that older patients, African-Americans and HIV/HCV coinfected patients respond less well (Feld and Hoofnagle, 2005; Fleckenstein, 2004). A correlation has been found between a small stretch of sequence within the NS5A region of the viral genome and therapeutic outcome. This region has been termed the interferon sensitivity-determining region (ISDR) because of its potential role in modulating the response to interferon therapy, and is said to do this by directly interacting with one or more cellular proteins associated with the interferon-mediated antiviral response (Katze *et al.*, 2002; Pawlotsky, 1999; Pawlotsky, 2003b; Pawlotsky *et al.*, 1998). Finally, current therapy is expensive, significantly limiting treatment to within developed countries, and displays a number of serious side effects in patients, including depression, fatigue, anaemia and neutropenia. Consequently, there is an immediate urgency for the development of new, more cost-effective treatments that target the virus more effectively, while causing as few adverse effects to the patient as possible.

Research into new means of anti-HCV therapy has focused on viral enzymes, host proteins involved in viral replication or cell entry, and the development of an effective vaccine. The problem faced by these efforts is the highly genetically heterogeneous nature of

the virus, along with its extraordinary replication pace and mutation rate. This virus has an incredible ability to quickly create variants of itself that can evade the immune system and escape the action of antiviral drugs. As a result, this characteristic has hampered the search for an effective vaccine or antiviral drugs. Propagation of HCV *in vitro* has been exceptionally difficult and only recently has a successful infectious cell-culture replicon system been developed (Lindenbach *et al.*, 2005; Wakita *et al.*, 2005; Zhong *et al.*, 2005), providing a powerful tool for the investigation of the viral life cycle and the discovery, assessment and development of new antiviral agents. It is fully developed only for a genotype 2 strain and this approach needs to be expanded to include more genotypes and subtypes in order to accommodate the broad genetic variability of clinical isolates. The replicon system also facilitates early analysis of viral resistance, cross-sensitivity studies and assessment of the efficacy of combination therapy and has been adapted for high-throughput screening of new drugs (Hao and Duggal, 2009). Progress in HCV research has also been hindered by the lack of an appropriate small animal model. This is because HCVcan only be transmitted to humans and chimpanzees. Developments have been made using a mouse model with a chimeric human liver (Mercer *et al.*, 2001; Meuleman *et al.*, 2005), but, although an extremely significant step, it is limited in its availability and applications. These slow, but significant advances in HCV research have made the search for anti-HCV drugs easier and produced some promising potential candidates for treatment.

While efficient preventative vaccines are available for HBV, efforts are still ongoing for the development of one for HCV. Vaccination trials against HCV in humans and chimpanzees were not initially successful, as re-infection was reported (Lai *et al.*, 1994; Prince *et al.*, 1992). The demonstration of the existence of a natural immunity to HCV, however, allows optimism that the development of an at least partially protective vaccine is achievable (Bassett *et al.*, 2001; Houghton and Abrignani, 2005; Weiner *et al.*, 2001). An affordable and globally available vaccine would be of utmost importance to reduce the significant numbers of new HCV infections per year.

The primary focus of research efforts for combating HCV involves the development of new specific small molecule inhibitors targeting stages of the virus life cycle, i.e. the viral NS genes in particular the NS3 serine protease and the NS5B RNA polymerase, which have been characterised at the biochemical and structural levels (de Bruijne *et al.*, 2009a; De Francesco and Carfi, 2007; Soriano *et al.*, 2009; Webster *et al.*, 2009). It should be noted that, although abundant structural data may be obtained by X-ray crystallography, the *in vivo* mechanism of

action and their precise role during replication are not fully understood, creating an obstacle in the development of the most efficient inhibitors.

To date, many inhibitors that have reached very promising stages of clinical trials have demonstrated dangerous side effects, halting their development, and so the search is still on for a drug that can be potent without causing toxicity. The other issue with small molecule inhibitors is that, when they are used in monotherapy, drug-resistant mutations rapidly develop and resistant strains are selected within the host, increasing the viral load again (Pawlotsky, 2003c; Thompson and McHutchison, 2009). Also, due to genetic differences, the efficacy of one inhibitor may differ greatly between genotypes, especially considering that these drugs are designed based on genotype-specific three-dimensional protein structures (Hinrichsen *et al.*, 2004; Kuntzen *et al.*, 2008a; Reiser *et al.*, 2005). This specificity suggests that drug regimens will be developed in a genotype-specific manner so that not all patients receive the same treatment. The success of future anti-HCV drugs depends on the fitness of resistant mutants and on the availability of suitable combination partners. It is likely that multiple drugs, targeting different proteins, will be developed and combined to provide the broadest spectrum of action on the different HCV genotypes and to avoid the rapid emergence of drug-resistant mutations. This will cause a significant change in HCV clinical management and crucial questions include what combinations will be most effective in which patients, whether HCV-targeted drugs will be sufficient for successful elimination of the virus, and whether patients can be cured without stimulation of the host immune system by either interferon, therapeutic vaccines or other immunomodulatory drugs.

The mechanisms underlying the ability of HCV to persist for the lifetime of the host are still open questions for the scientific community, and they need to be completely understood in order to aid anti-HCV research further. As it stands, huge efforts are underway for drugs that are under development to safely reach the public and meet the requirements of the urgency to prevent the global morbidity and mortality rates predicted by the current prevalence and incidence of HCV infection. Even though it is possible that a cure may one day be found for hepatitis C, the best strategy to stop this silent epidemic is prevention. In order to reduce risks of infection, to catch new infections early, and to minimise the risks of transmission, prevention efforts should focus on safer blood supplies and injection practises in developing countries, reducing the incidence of initiation of injecting drug abuse, on reducing risks among current users, introducing routine testing within high-risk cohorts, raising general

public awareness, and identifying, counselling and providing clinical management for already infected individuals.

1.8. Aims and outline of the thesis

The main aim of this thesis is to contribute to the current knowledge on HCV genetic diversity, molecular evolution and epidemiology, with particular reference to Cyprus. To do this, the project is divided into two large sections, which differ in scope and methodology.

In the first part of the thesis the aim is to portray a representative picture of the current molecular epidemiology and genotype distribution of HCV infection in Cyprus for the first time. The objectives include the design of a reverse transcription (RT)-PCR and sequencing protocol that has the ability to efficiently amplify strains of all HCV genotypes in two genomic regions, and to apply this protocol to samples available from the general population and from specific high-risk cohorts. Specifically, the groups to be investigated include the general HCV-infected population, intravenous drug users seeking therapy in rehabilitation centres, a cohort of incarcerated individuals and patients coinfected with HIV-1. The different cohorts were sampled and studied sequentially, through collaborations established between the laboratory of Biotechnology and Molecular Virology and various clinicians, the Anti-drugs Council, the National Monitoring Centre for Drugs and Drug Addiction and clinicians in the State Correctional Facilities. Cyprus is a small country with a population of approximately 800,000, but unique with respect to its geographical position, as it lies between Europe, Africa, and Asia. It is very close to North Africa, where the prevalence of HCV is the highest in the world. It has a high rate of influx of tourists and political and economic immigrants from Eastern Europe and countries of the former Soviet Union, Africa and Southeast Asia. Also, as it is now a member state of the European Union, entry into Cyprus is easy, thus facilitating the introduction of new infectious diseases. As such, it is expected that a certain heterogeneity will be observed in the genotype distribution of HCV, as seen for HIV (Kousiappa *et al.*, 2009a; Kousiappa *et al.*, 2009b). In this section, the experimental design will be presented and discussed, along with the patient characteristics of each group, the genotype distribution within and among the cohorts, and the relationships of the identified strains with each other and with strains from international databases, through phylogenetic analysis and interpretation of clusters. The purpose of this is to establish the degree of phylogenetic heterogeneity of HCV on the island and to discuss the possible sources of HCV import from knowledge on global molecular epidemiology, which has never before been

attempted. Additionally, the intention was also to submit the identified sequences to public databases in order to make them available to the scientific community for further epidemiological studies.

The second part focuses on specific strains identified from the molecular epidemiology study. The aim was to select rare isolates and analyse them in more detail by near-full genome sequencing. The objective was to design and develop strain-specific protocols for the specific amplification and sequencing of genome fragments of particular isolates. Specifically, recombinant strains, which are uncommon for HCV, and novel strains which have not yet been classified into subtypes have been presented and submitted to sequence databases in order to contribute to current knowledge of HCV genetic variability and classification.

2. Part I: Molecular Epidemiology of HCV in Cyprus

2.1. Specific objectives

The objectives in this section of the thesis are to study the molecular epidemiology of HCV infection in Cyprus and uncover the genotype distribution in various risk groups. Specifically, the aims were to establish collaborative networks to facilitate sampling, to design an RT-PCR and sequencing protocol for amplification of all HCV genotypes in two genomic regions, and to carry out phylogenetic analysis of the strains in order to investigate the current sequence diversity and the points of introduction of the virus onto the island. The groups were investigated in a consecutive manner, and included HCV-infected individuals from the general population seeking medical care, intravenous drug users in rehabilitation centres, individuals from the state correctional facilities undergoing routine HCV testing, and samples from another project focusing on the molecular epidemiology of HIV in Cyprus, which were diagnosed positive for HCV in routine diagnostic tests. All sequences have been made publicly available for publication in the GenBank international sequence database.

2.2. Methods

2.2.1. Study subjects

2.2.1.1. General population (2005-2009)

From 2005 to 2009 blood samples were obtained from 111 chronically infected HCV patients aged 18-84, who were attending private clinics and public hospitals in the Nicosia, Larnaca, Limassol, and Paphos districts. This was made possible by setting up a long term collaborative network (the Cyprus HCV Network) with hepatologists and gastroenterologists in these institutions. All patients tested positive for HCV antibodies by a second-generation immunoassay (INNO-LiPA), and for HCV RNA by diagnostic RT-PCR (COBAS Amplicor, Roche Diagnostics, Branchburg, NJ, USA). These tests were carried out as part of routine tests and were done externally and independently of this research. All patients included in this study gave their informed consent, and clinical and epidemiological data were collected by use of questionnaires. Each questionnaire and corresponding sample was given a laboratory code to protect the identity of each participant. Blood sampling for this study was carried out by qualified personnel at each hospital or clinic.

2.2.1.2. Intravenous drug users (2008)

Following the establishment of the ongoing collection of HCV samples from the general population of Cyprus, a separate collaborative network was set up with the Cyprus National Monitoring Centre for Drugs and Drug Addiction, the Cyprus Anti-drugs Council and all detoxification centres on the island, in order to carry out sampling in a more specialised high-risk cohort. The study subjects were consenting intravenous drug users seeking therapy in detoxification centres around the island at the time of sample collection. Two rounds of sampling were carried out in July and November 2008. The rehabilitation centres where the sampling took place were as follows: Aghia Skepi, a long-term inpatient therapeutic community for withdrawal from substance abuse and social rehabilitation; Ghephyra, an outpatient substitution treatment program, for users who have failed with other therapeutic programs and buprenorphine is administered in parallel to psychosocial intervention and psychological and medical aid; Anosis, a long-term inpatient detoxification program, for abstinence from drug use and safe management of withdrawal symptoms through administration of analgesics and psychosocial support; Ploeghos, an outpatient psychosocial treatment program, for users who have failed with other therapeutic programs and buprenorphine is administered in parallel to psychosocial intervention and psychological and medical aid; Veresie Clinic, a clinic offering detoxification services through psychological and medical services, and where naltrexone implant procedures are performed. An informed consent form was signed by each subject, a questionnaire was filled in with an interviewer from the Cyprus Monitoring Centre for Drugs and Drug Addiction, and blood samples were taken by qualified personnel as with the subjects from the general population. All samples and questionnaires were coded at random with a laboratory identifier number so as not to reveal the personal identity of the study subjects. Immunoassays were performed on all samples for detection of HCV, HBV, HIV-1 and HIV-2 antibodies, using the AxSYM anti-HCV, HIV and HBsAg systems (Abbott Diagnostics, Chicago, IL). These tests were done externally at a clinical laboratory certified for carrying out routine diagnostics assays. Forty one individuals in total were included in this study, representing 87% of intravenous drug users requesting detoxification support at the time of sampling. This work has obtained bioethical approval from the Cyprus National Bioethics Committee (EEBK/EQ/2008/15).

2.2.1.3. Incarcerated population (2009-2010)

For the purposes of routine diagnostic testing, twenty five blood samples were collected from persons serving a prison sentence at the Cyprus state correctional facilities, between June 2009 and January 2010, in collaboration with the clinician in charge. Blood samples were collected by qualified personnel at the penitentiary clinic and were safely transported to the laboratory of Biotechnology and Molecular Virology for processing. Sixteen of these were tested PCR-positive for HCV and were therefore selected retrospectively to be included in the study of the molecular epidemiology of the HCV infection in Cyprus, as a high-risk cohort. The sixteen individuals were asked for written consent to participate and for their samples to be included in this project. Seven of these study subjects were available to fill out a questionnaire and epidemiological data were thus obtained. Each sample and questionnaire was coded with a random laboratory identifier code for protection of the identity of the study subjects.

2.2.1.4. HIV/HCV coinfected cohort (2005-2009)

The final high risk cohort looked at in this study were HIV-1 patients coinfected with HCV. These samples were chosen from a bank of HIV-1 samples maintained by the Laboratory of Biotechnology and Molecular virology through its ongoing collaboration with the Gregorios AIDS Clinic, which is the National Reference Clinic for HIV in Cyprus. HIV-1 seropositive samples were collected from consenting patients and plasma is stored for the ongoing project of the molecular epidemiology and genotypic drug resistance of HIV-1 in Cyprus (Kousiappa *et al.*, 2009a). From the clinical data available, all samples known to be coinfected with HCV were selected for this study (14 samples) and previously stored aliquots of plasma were used. The samples are numbered with a laboratory identifier code and clinical and epidemiological data previously collected for the HIV-1 project was accessed by using these codes.

2.2.2. Sample collection and processing

Blood (4-5 ml) was collected from each participant in a BD Vacutainer® PPT™ (Becton Dickinson and Co., Franklin Lakes, NJ) tube and transported safely to the Laboratory of Biotechnology and Molecular Virology. Plasma was isolated after centrifugation at 1,100 RCF (relative centrifugal force) for 10 min in an Eppendorf Centrifuge 5810 R (Eppendorf), and stored in aliquots at -70^{o}C until it was used for RNA extraction. Viral RNA was extracted

from 200 μl plasma using the QIAmp® UltraSens® Virus kit (Qiagen, Venlo, the Netherlands) following the manufacturer's instructions. All potentially infectious material was handled in a bio-safety level 3 area following appropriate health and safety measures.

2.2.3. Experimental design

All HCV RNA samples were amplified by a nested RT-PCR. Partial sequences from the Core-E1 and NS5B regions of the HCV genome were targeted. For each region, degenerate forward and reverse primers were designed to amplify all HCV genotypes (Table 1) by manually studying multiple alignments of HCV sequences covering all subtypes available on the Los Alamos HCV sequence database (http://hcv.lanl.org), to perform inner and outer PCR. All PCR-positive samples were sequenced using the inner forward and reverse primers. The experimental design is described schematically in Figure 3.

Table 1. Primers for amplification and sequencing of the Core-E1 and NS5B regions.

Name	Type	Polarity	Sequence (5'-3') [a]	Position [b]
Core-E1				
735	Outer	Forward	GACCTCATGGGGTACATYCCBSTCGTHGG	735-763
834	Inner	Forward	GCAACAGGGAATYTDCCYGGTTGCTCYTTYTC	834-865
1318	Inner	Reverse	CAGTTCATCATCATRTCCCAWGCCATNCGRTGDCC	1,284-1,318
1324	Outer	Reverse	GGBGACCARTTYAKCATCATRTCCCAWGCC	1,295-1,324
NS5B				
8172	Outer	Forward	TAYGGRTTCCARTACTCNCCHGVRCAGCGGGT	8,172-8,203
8244	Inner	Forward	ATGGGBTTYKCRTATGAYACCCGHTGYTTTGA	8,244-8,275
8713	Inner	Reverse	GABACRTTKGAGGARCADGATGTTATNARCTC	8,682-8,713
8821	Outer	Reverse	GARTTGACWGGRGWGTGTCKDRCTGTYTCCCA	8,790-8,821

[a] Degenerate positions are shown with IUB base codes (R: A or G; W: A or T; S: G or C; K: G or T; Y: C or T; B: C, G or T; D: A, G or T; H: A, C or T; V: A, C or G; N: A, C, G or T).
[b] Positions numbered according to strain H77 (GenBank Acc. No. NC_004102).

2.2.4. Reverse transcription nested PCR and sequencing

15 μl of the RNA was used in a one-step RT-PCR using Superscript™ III One-Step RT-PCR Platinum Taq HiFi (Invitrogen, Carlsbad, CA), following a heat-shock step at 70 ^{o}C

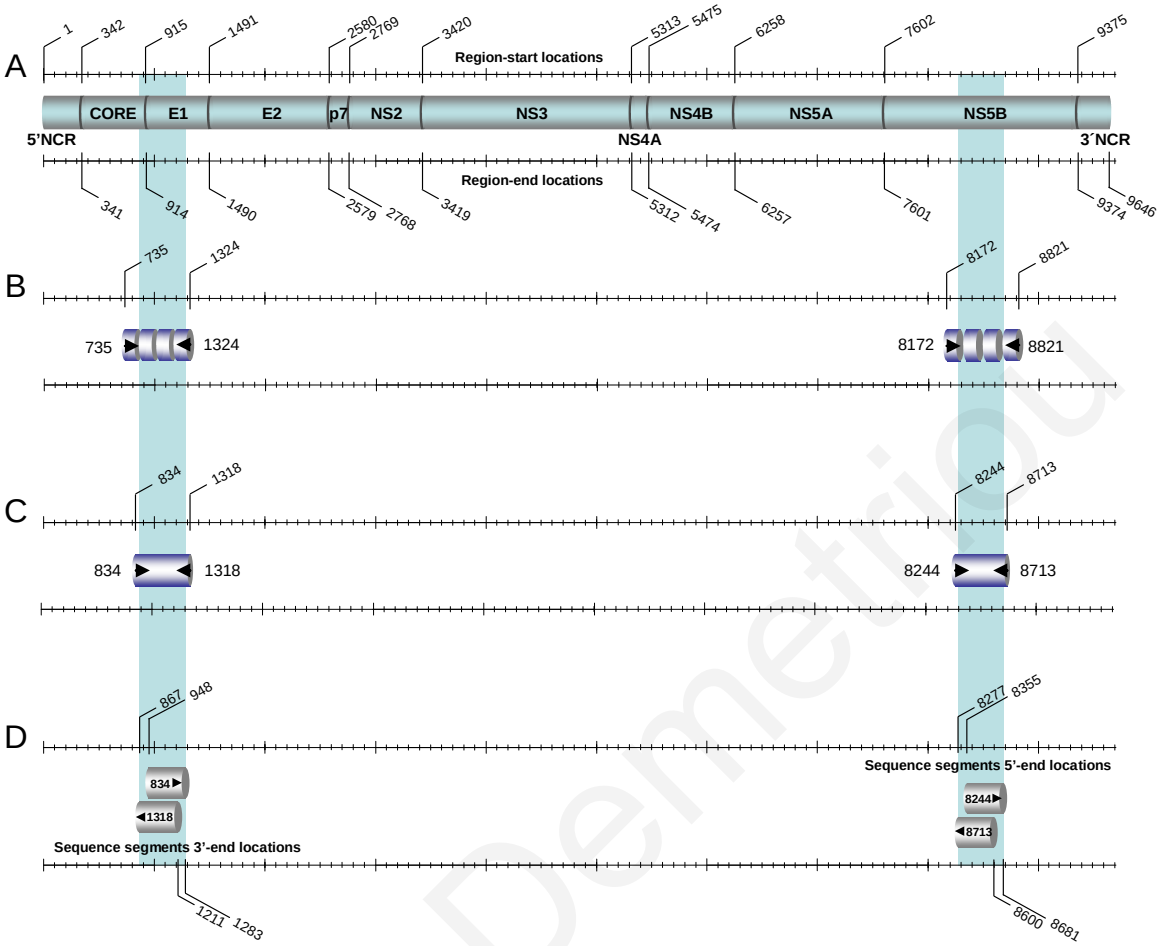

Figure 3. Schematic diagram of the experimental design of PCR and sequencing for the two regions studied in the molecular epidemiology. (A) The HCV genome coordinate map numbered according to strain H77 (Acc. No. NC_004102); (B) the outer PCR product locations with numbers designating the names and 5' ends of the corresponding outer PCR primers; (C) the inner PCR product locations with numbers designating the names and 5' ends of the corresponding inner PCR primers; (D) the sequencing product locations with the numbers on the scale indicating the start and end of each reading, and the numbers and arrows in the cylinders indicating the sequence primer name and directionality.

for 20 s to denature the RNA secondary structure. The RT-PCR was performed in a 50 µl reaction with 20 pmol each of the outer primers (Table 1). Reverse transcription was carried out at 55°C for 60 min, and then 94°C for 2 min. This was followed by 43 cycles of 94 °C for 15 s, 57 °C for 30 s and 68 °C for 1 min per kilobase of product, and a final elongation step at 68 °C for 5 min. A nested PCR was performed using 3 µl of the first round PCR product with 40 pmol each of the inner PCR primers (Table 1), using Platinum® PCR SuperMix (Invitrogen, Carlsbad, CA, USA) in a 50 µl reaction. Cycling conditions were as follows: 94 °C for 2 min, 40 cycles of 94°C for 20 s, 30 s at the appropriate annealing temperature and 72°C for 1 min per kilobase of product, and a final elongation step at 72°C for 5 min. The

annealing temperatures for the secondary Core-E1 and NS5B PCR were 56 ^{o}C and 53 ^{o}C, respectively. PCR amplification was confirmed by visualisation of bands with ethidium bromide staining on a 2% agarose gel and PCR products were purified using the QIAquick ® PCR Purification Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The concentration of each PCR product was measured at 260 nm on the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Cycle sequencing reactions were carried out on the amplicons of all PCR-positive samples bidirectionally using the inner forward and reverse PCR primers for each region (Table 1) by means of the BigDye® Terminator system v3.1 (Applied Biosystems, Foster City, CA). The reactions were performed at 94 ^{o}C for 1 min, followed by 25 cycles of 96 ^{o}C for 10 s, the appropriate annealing temperature for 5 s, and 60 ^{o}C for 4 min. The annealing temperatures were 57 ^{o}C for primer 834, 57 ^{o}C for primer 1318, 56 ^{o}C for 8244, and 53 ^{o}C for 8713. The products were purified using the DyeEx spin kits (Qiagen) and sequenced directly on the ABI 3300 Genetic Analyser (Applied Biosystems). The resulting readings were analysed with the Sequencing Analysis Software v5.2 (Applied Biosystems) and checked manually. The nucleotide sequences obtained for both the Core-E1 region (417 bp, positions 867-1283 according to strain H77) and NS5B region (405 bp, positions 8277-8681 according to strain H77) were aligned with the H77 nucleotide reference sequence using the CLUSTALX 1.83 alignment software with default parameters (Thompson *et al.*, 1997).

2.2.5. Subtyping and phylogenetic analysis

Subtyping based on the sequence obtained for each region was performed using the Oxford HCV Subtyping Tool v1.0 (de Oliveira *et al.*, 2005), after which the aligned sequences were compared to reference strains of known subtypes derived from the Los Alamos HCV database (Kuiken *et al.*, 2005) using the neighbour-joining method (Saitou and Nei, 1987) in MEGA version 4 (Tamura *et al.*, 2007). Pair-wise distance matrices were generated using the Kimura-2-parameter distance estimation approach (Kimura, 1980). The reliability of the phylogenetic clustering was evaluated using bootstrap analysis with 1,000 replicates (Felsenstein, 1985). Bootstrap values above 70% were considered sufficient for subtype assignment. Clusters were determined by an arbitrary threshold of genetic distance < 5% (Lewis *et al.*, 2008).

For the purpose of a deeper investigation into possible phylogenetic relationships between the dataset obtained from the intravenous drug users cohort and other HCV strains,

each sequence was uploaded individually into the HCVBLAST tool on the Los Alamos HCV database website (http://hcv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html) to recover the most closely related sequences available in the database (>95% sequence similarity, as defined by the percentage of identical residues per position along the given length of sequence). For a deeper and more sophisticated analysis, phylogenies for these sequences were constructed with MrBayes (Huelsenbeck and Ronquist, 2001) using a general time-reversible (GTR) model of nucleotide substitution with a proportion of invariant sites (i) and gamma distribution of rates (G). The Monte-Carlo Markov Chain search was run for $5x10^6$ generations, with trees sampled every 100^{th} generation (with a burn-in of 50%) and a posterior consensus tree generated (from 25,000 trees). From this consensus tree, the posterior probability of nodes was used as phylogenetic support for each transmission cluster group. Cluster group size was determined using nodes with a posterior probability of 1.

2.2.6. Reference strains

The GenBank accession numbers of the reference strains used in each phylogenetic analysis are shown in the respective figures of the neighbour-joining trees in the results section, apart from the general analysis of all Cypriot strains, for which the accession numbers are listed here. For both the Core-E1 and NS5B regions the reference sequences used were: 1a, AF511950, EF407419 and NC_004102; 1b, AY587016, D11355 and EF032892; 1c, AY051292 and D14853; 2a, AB047639, AY746460 and D00944; 2b, AB030907, AF238486 and D10988; 2c, D50409; 2k, AB031663; 2k/1b, AY587845; 3a, AF046866, D17763 and X76918; 3b, D49374; 3k, D63821; 4a, DQ418788 and Y11604; 4d, DQ418786 and DQ516083; 5a, AF064490 and Y13184; 6a, AY859526, DQ480513 and Y12083; 6b, D84262; 6c, EF424629; 7a, EF108306.

2.2.7. Statistical analysis

To test for a statistically significant correlation between the PCR results and demographic and clinical variables from the general population samples, the χ^2-test for categorical variables and the *t*-test for continuous variables were used. In the intravenous drug user cohort, variables were compared between hepatitis C positive and negative individuals using the Mann-Whitney test for continuous variables and χ^2-test or Fisher exact test where appropriate.

2.2.8. Nucleotide sequence accession numbers

The accession numbers for sequences obtained in this study and submitted to GenBank are EU684591-737, HQ537010-2 & HQ537033-6 for the general population sequences, GQ332540-65 for the intravenous drug users' sequences, HQ537013-25 & HQ537037-52 for the sequences of the incarcerated population, and HQ537026-32 & HQ537053-9 for the sequences of the HIV/HCV coinfected group.

2.3. Results and Discussion

2.3.1. PCR fidelity

On an analysis performed on 107 samples from the general population, 77 samples were PCR-positive for Core-E1 and 70 for NS5B, presenting 73.8% and 71.0% PCR success rates, respectively. Statistical analysis to determine the association between PCR result and whether the patients were on therapy revealed a statistically significant more frequent negative PCR result for patients on therapy with *P*-values < 0.016. Fifty-six patients (52.3%) were on interferon therapy when blood was taken, and 51 patients (47.7%) were also taking ribavirin. Forty-nine patients (45.8%) were not on therapy at the time blood was taken. Considering the two groups separately, from the patients on interferon therapy, 35 exhibited PCR-positive results and 21 showed negative results, compared to 44 positive results and 7 negative results from the patients not on interferon (*P*=0.008). Of the patients also taking ribavirin, 32 had positive PCR results and 19 had negative results compared to 47 positive and 9 negative samples from patients not taking ribavirin (*P*=0.016). Furthermore, the correlation between PCR result and viral load showed that samples from patients with a detectable viral load (more than 135 copies/ml) exhibited more frequently a positive PCR result than those with undetectable viral load with a *P*-value of 0.002. From the patients with undetectable viral load at the time blood was taken, 6 had positive PCR results and 13 had negative results, compared to patients with detectable viral load, of which 58 were positive and 21 negative (*P*=0.002; patients with unknown viral load were excluded from this analysis).

An analysis was also performed on the intravenous drug user samples, whose antibody test results were available. Of the 20 samples found to be HCV positive in the AxSYM anti-HCV system, only 14 were PCR positive in the Core-E1 region and 12 of these in the NS5B region also. PCR success rate was determined by comparing PCR results (positive or negative) with the results from AxSYM, which gives a value for the sample/cut-off rate ratio (S/CO) for

each sample, where a value above 1 is considered positive. The S/CO values for the positive samples in this study were 4-124. A Mann-Whitney test showed that the PCR positive samples were those with significantly higher S/CO values than the PCR negative samples for both regions tested, as seen in Table 3 (P=0.001). It is known that chronic HCV patients generally do have higher anti-HCV antibody titres than convalescent individuals (Houghton, 2009).

Therefore, viral RNA extraction from blood plasma, RT-PCR and sequencing of the Core-E1 and NS5B regions were designed and successfully applied for genotyping of HCV strains, enabling detection of at least genotypes 1-5, and the determination of the genetic heterogeneity of HCV in this cohort of patients from Cyprus. The RT-PCR assay design is considered successful as over 70% of samples were PCR-positive in both regions and those that were not belonged mainly to patients receiving therapy and/or with low viral load with statistical significance (low P-values).

2.3.2. Patient characteristics

2.3.2.1. General population (2005-2009)

The study group consisted of 111 HCV seropositive patients between the ages of 18 and 84, from two private gastroenterology clinics and all public hospitals in Cyprus. A list of the patient demographic characteristics is seen in Table 2. The epidemiological features of the study subjects varied, as 53 (49.1%) patients were Cypriots and 37 (34.3%) being from countries of the former Soviet Union (Russia, Georgia, Moldova, Belarus and Ukraine), but the group also included patients from other countries, such as Greece, Britain, Pakistan, Italy, Egypt, Iran, Romania, and Bulgaria.

Thirty-nine patients (35.1%) had a history of transfusion with blood products, six (5.4%) stated they were intravenous drug users, ten (9.0%) traced infection to dental or surgical procedures, and 52 (46.8%) did not know the source of infection. The route of transmission for a large percentage of patients was unknown, mainly because this virus can be carried for many years without diagnosis and because safety measures for the prevention of HCV transmission in hospitals and screening of donated blood and blood products were not carried out until the early 1990s. Also, considering the patients' diverse nationalities, it is worth noting that unsafe medical practice still occurs in some developing countries. The patients investigated in this group were therefore epidemiologically diverse, showing a wide range of ages, countries of origin, and routes of transmission for HCV.

Table 2. Characteristics of the general population cohort.

Characteristics	Patients (N=111)	
Gender (%) [a]		
Male	55 (50)	
Female	53 (48)	
Age (years) (%) [a]		
Median (IQR)	40 (32-57)	
18-29	23 (21)	
30-39	33 (31)	
40-49	11 (10)	
50-59	22 (20)	
60 and older	19 (18)	
Region of origin (%) [b]		
Cyprus	53 (50)	
Russia	15 (14)	
Georgia	14 (13)	
Greece	5 (5)	
Moldova	4 (4)	
Bulgaria	3 (3)	
Ukraine	3 (3)	
Egypt	3 (3)	
Other - Europe	3 (3)	
Other - Asia	3 (3)	
Route of transmission (%)		
Blood transfusion	39 (35)	
Surgical procedure	6 (5)	
Dental procedure	4 (4)	
Intravenous drug use	6 (5)	
Tattoo	2 (2)	
Sexual transmission	1 (1)	
Other / unknown	53 (48)	
Use of medication (%)		
Interferon – ribavirin	59 (53)	
No medication	52 (47)	
Plasma HCV-RNA (log copies/ml)		
Median (IQR)	5.5 (2.5-6.2)	
	Core-E1	**NS5B**
RT-PCR (%)		
Positive	82 (74)	80 (72)
Negative	29 (26)	31 (28)
Genotype (%)		
1	52 (47)	49 (44)
2	3 (3)	1 (1)
3	15 (14)	15 (14)
4	9 (8)	8 (7)
5	1 (1)	1 (1)
Unknown	31 (28)	37 (33)

IQR, interquartile range.
[a] Information available for 108 subjects.
[b] Information available for 107 subjects.

2.3.2.2. Intravenous drug users (2008)

In this part of the study, the prevalence and molecular epidemiology of HCV among intravenous drug users in Cyprus seeking rehabilitation was investigated. Forty one consenting individuals took part, providing demographic and epidemiological characteristics as well as information on risk groups and drug use behaviour (Table 3). The majority of the participants were male (85%), young, with a median age of 27 years (interquartile range 25-31) and of Cypriot nationality (64%). Other nationalities included either Greek or Eastern European, demonstrating the influx of young people from Eastern Europe to the island and its strong association with Greece. There was no statistical significance between the distribution of age or nationality between the HCV positive and negative individuals (P=0.61 and 0.30 respectively). The results of the antibody testing showed that none of the intravenous drug users who took part in the study were infected with HIV or HBV at the time of sampling, surprisingly signifying a zero prevalence of these infections in this population. However, 20 individuals of the 41 tested were positive for HCV antibodies, giving a prevalence of 49%, which is comparable to or lower than known prevalence rates in intravenous drug users abroad (Shapatava *et al.*, 2006; Shepard *et al.*, 2005; Sutton *et al.*, 2008; Tan *et al.*, 2008; van de Laar *et al.*, 2005).

In this population of highly experienced intravenous drug users the main statistically significant factor predicting HCV was the duration of injection drug use, which, with a median duration of 10 years (interquartile range 5-14) was significantly higher in HCV-positive individuals as compared to a value of 6 years (interquartile range 1.5-8.5) in the HCV-negative group (P=0.02). Duration of injecting drug use practices is a factor known to be a predictor of higher HCV prevalence (Hagan *et al.*, 2007; Shepard *et al.*, 2005; Sweeting *et al.*, 2009). The association of the duration of intravenous drug use with the prevalence of HCV infection is seen in Figure 4. Another risk factor significantly different between the HCV positive and negative groups was a history of prison sentence. HCV-positive users reported having served a prison sentence more frequently than those who were negative (P=0.01), highlighting the occurrence of unsafe injecting practices in correctional facilities. All subjects reported to have practiced unprotected sex, but only two stated MSM behaviour, given the recent reports regarding MSM being a route of transmission (van de Laar *et al.*, 2007). Interestingly, there were no significant differences between the HCV-positive and negative groups with regards to having shared injecting equipment or having easy access to clean syringes, or other stated risk behaviours.

Table 3. Characteristics of the intravenous drug user cohort.

Characteristics [a]	**Subjects (N=41)**	**HCV +ve**	**HCV -ve**
Age (years) (%)			
Median (IQR)	27 (25-31)	27 (24-34)	27 (25-29)
Gender - median (%)			
Male	35 (85)	16	19
Female	6 (15)	4	2
Nationality - median (%)			
Cypriot	26 (64)	12	14
Greek	8 (21)	4	4
Russian	2 (5)	2	0
Bulgarian	1 (3)	1	0
Georgian	1 (3)	1	0
Romanian	1 (3)	0	1
Not stated	2 (3)	1	1
Risk factors - median (%)			
Age at first injection (years) (IQR)	20 (17-23)	19.5 (16-21)	21 (18-25)
Duration of injecting drug use (years) (IQR)		10 (5-14)	6 (2-9)
Served a prison sentence	11 (28)	9	2
Shared injecting equipment	27 (68)	13	14
Can find sterile equipment easily	26 (65)	12	14
History of blood transfusion	3 (8)	3	0
History of surgical procedure	24 (60)	12	12
Tattoos	25 (63)	12	13
Have used syringes/needles abroad	10 (25)	7	3
Sexual practices - median (%)			
Heterosexual	35 (92)	17	18
Homosexual	1 (3)	1	0
Bisexual	2 (5)	0	2
Unprotected sexual practices	41 (100)	20	21
RT-PCR [b]	*Core-E1*	*NS5B*	
Positive	14	12	
Negative	6	8	
AxSYM anti-HCV S/CO values (IQR) [c]	*Core-E1*	*NS5B*	
PCR positive	86.8 (47.8-95.7)	88.1 (77.8-95.9)	
PCR negative	8.8 (4.6-25.7)	14.6 (5.5-45.6)	
Genotype [d]	*Core-E1*	*NS5B*	
1b	6	4	
3a	8	8	

[a] IQR, interquartile range

[b] Only performed on samples which were HCV seropositive

[c] AxSYM anti-HCVsystem, the HCV antibody test used in this study; S/CO, signal/cutoff rate ratio is the result obtained from the HCV antibody test

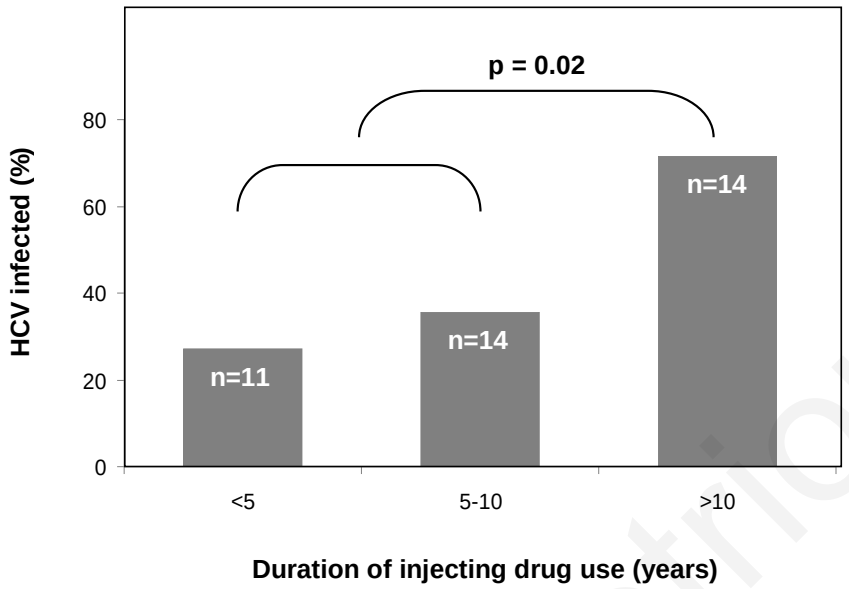

Figure 4. Bar chart indicating the percentage prevalence of HCV seropositive injecting drug users in three groups of different duration of practise. In the group of individuals who have been users for up to 5 years, 3 out of 11 are anti-HCV positive (prevalence is 27%); within the individuals who have practiced IDU for 5-10 years, 5 out of 14 are anti-HCV positive (prevalence is 36%); within the individuals who have been practising IDU for over 10 years, 10 out of 14 are positive (prevalence is 71%). Intravenous drug users who have been practising for more than 10 years are significantly at higher risk of being HCV seropositive than short-term users (p=0.02).

2.3.2.3. Incarcerated population (2009-2010)

Following the results of the injecting drug users' population, and the fact that serving a prison sentence was shown to be a significant risk factor for HCV infection, the study of the molecular epidemiology of HCV in Cyprus proceeded to the next step: a total of 25 samples from individuals serving a prison sentence had been previously taken for standard diagnostic purposes and any samples testing PCR-positive for HCV were retrospectively added to this study after written consent from the study subjects (16 individuals). Interviews were carried out with as many study subjects as possible and epidemiological and risk-behaviour details were collected from seven individuals. The participants were predominantly male (6, 85%), aged between 22-41 years, four (57%) were Cypriot, and three (43%) were of other nationalities (Greek, Georgian and Portuguese). All individuals who had been previously tested for coinfection with HBV and HIV (6, 85%) stated that they were tested negative for either infection. Three subjects had served multiple prison sentences. All interviewed persons stated that they had practiced intravenous drug use, six of which in Cyprus, and three had also practiced injecting drug use abroad. The median stated duration of injecting drug use was 5 years (interquartile range, 2-6 years). Regarding other possible routes of transmission and

high-risk behaviours, none of the study subjects had ever had a blood transfusion, six (85%) had tattoos, all had practised unsafe heterosexual contact, and one participant believed that he had been infected with HCV during a fight within the correctional facilities. These risk behaviours are representative of the HCV risk factors associated with prisons (Babudieri *et al.*, 2005; Miller *et al.*, 2009a; Zalumas and Rose, 2003).

2.3.2.4. HIV/HCV coinfected cohort (2005-2009)

The final risk group investigated here are HIV/HCV coinfected subjects, who are considered to arise from two types of risk groups: injecting drug users and MSM (Kim and Chung, 2009; Re *et al.*, 2008; Tan *et al.*, 2008; Urbanus *et al.*, 2009; van de Laar *et al.*, 2009). From the 251 samples collected from HIV-1 seropositive patients from 2005-2009 for the study of the molecular epidemiology of HIV-1 in Cyprus (Kousiappa *et al.*, 2009b), 14 were known to have HCV coinfection, from the results of routine diagnostic tests. All coinfected samples were subsequently used in this study, six of which (43%) were on antiretroviral therapy at the time of sampling. This cohort consisted predominantly of male subjects (10, 71%), with a median age of 34 years (interquartile range 32-45). The subjects displayed a variety of nationalities, with five (36%) being from Cyprus, four (29%) from Georgia, and one patient (7%) each from Bulgaria, Latvia, Italy, Britain and the USA. The stated routes of HIV-1 transmission for these patients was heterosexual contact for nine (64%), homosexual contact for two (MSM) (14%), intravenous drug use practises for two (14%), and one (7%) was a haemophiliac. No further specific details, however, were available regarding HCV transmission, although it is likely that the routes of transmission for the two viruses were the same in many cases. Of all the above, the most improbable route of transmission for HCV is the heterosexual contact, in which cases HCV may have been acquired through another unknown route. All patients appeared epidemiologically unlinked, apart from one male and one female subject from Georgia, who were partners at the time of sampling (Kousiappa *et al.*, 2009b).

2.3.3. Genotype distribution and phylogenetic analysis

2.3.3.1. General population (2005-2009)

The HCV genotype distribution and sequence variability were investigated in a group of hepatitis C patients from the general population in Cyprus who were seeking medical care

in public hospitals and private clinics between 2005 and 2009. The phylogenetic analysis was performed on two partial genomic regions, Core-E1 and NS5B, and the resulting trees are seen in Figure 5 (the samples from the general population are shown as yellow circles). HCV genotype 1 was the most frequent (47.7% in the Core-E1 region), followed by genotypes 3, 4, 2, and 5 (12.1%, 8.4%, 2.8%, and 0.9%, respectively in the Core-E1 region). There is concordance between the trees for all samples that were positive for both regions, except for two strains from Georgian patients (designated with asterisks in Figure 5). These strains were identified as subtype 2k in the Core-E1 region and as 1b in the NS5B region. They therefore appear to be 2k/1b recombinants, as they cluster together with the St. Petersburg 2k/1b recombinant strain in both trees (GenBank Acc. No. AY587845). The St. Petersburg strain is the reference strain in the trees closest to the isolates indicated with asterisks. However, near-full genome analysis is required to confirm this and is described in the second part of this study.

Within genotype 1, 37 general population recovered sequences were classified as subtype 1b. Two additional strains were also classified as subtype 1b in the Core-E1 region, but were PCR-negative in the NS5B region and were therefore not included in the NS5B analysis. Twenty-two of these strains are from patients originating from Cyprus, six from Georgia, five from Russia, three from Moldova, and one each from Greece, Ukraine, and Romania. One strain in particular, from a Greek patient, had a 3-nucleotide insertion in the Core-E1 sequence between positions 1044 and 1045 (positions according to strain H77, GenBank accession no. NC_004102), resulting in an additional amino-acid in the E1 protein sequence. The epidemiological origin or significance of this insertion is not currently known. Ten strains grouped within the 1a subtype in both regions with bootstrap values higher than 90. Three further isolates were also classified as subtype 1a in the Core-E1 region but were PCR-negative in the NS5B region and could hence not be analyses in this region. Seven are from Cypriot patients, two from Georgian patients and one each from a British, an Iranian, an Italian, and a Greek patient. One strain from a Cypriot patient did not cluster with any common genotype 1 subtypes using the standard reference sequences, and so, separate phylogenetic trees were built for both genomic regions using reference sequences from all assigned genotype 1 subtypes available within these regions in the HCV sequence database (Figure 6A & B). However, this strain still did not group with any known subtype. To investigate this strain further, a BLAST analysis was performed using the HCV BLAST tool

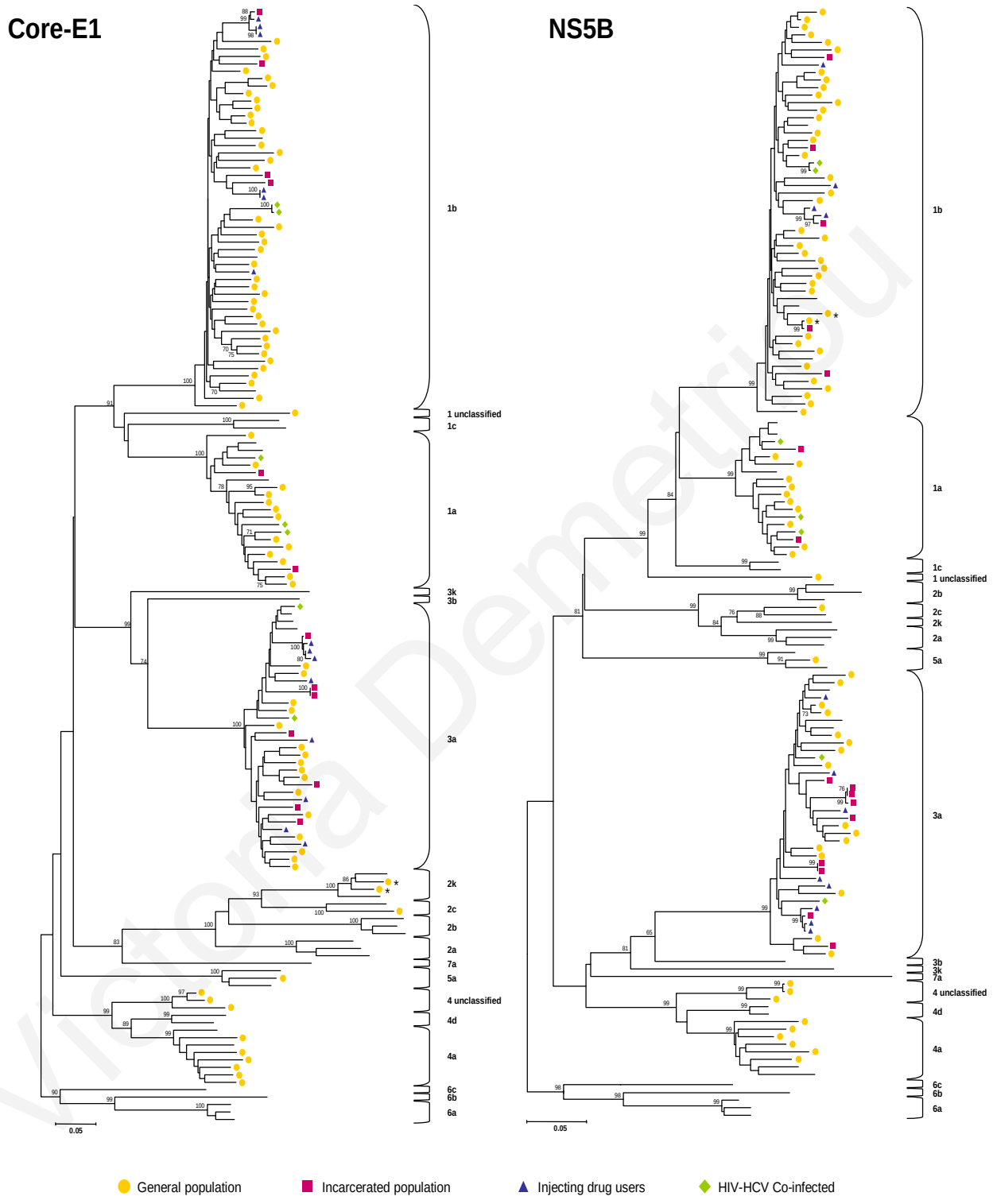

Figure 5. Neighbour-joining trees showing the genotype distribution of the HCV samples described in this study in the Core-E1 and NS5B genomic regions. HCV subtypes are designated with brackets on the right of each tree. The samples are denoted with different symbols for each cohort, as indicated below the trees. The genotype 6 group of strains were used to root the tree. The numbers at branch nodes are percentage bootstrap support for 1,000 replicates, and only values above 70 are shown. The genetic distance between any two sequences is denoted by the scale at the lower left side of each tree.

on the HCV Sequence database website, but no sequence on the database gave a genetic similarity higher than 74% or 82% in the Core-E1 and NS5B regions, respectively, which is not high enough similarity to signify subtype grouping (Simmonds *et al.*, 2005). This strain is hence labelled unclassified and is further analysed by near-full genome sequencing in the second part of this thesis.

For genotype 2, a single strain from a Cypriot patient clustered with subtype 2c in the Core-E1 and NS5B trees with bootstrap values of 98 and 83, respectively. This was isolated from a Cypriot patient who was born in Argentina and was infected there by a blood transfusion in the late 1970s. According to a recent study, 2c is a subtype found in high prevalence in central Argentina (Re *et al.*, 2007). However, it is unknown whether this was the case at the time this patient became infected, nor is it known where in Argentina the blood transfusion took place.

Also the two putative recombinant strains mentioned earlier clustered with the 2k subtype in the Core-E1 tree with a bootstrap of 99, but with the 1b subtype in the NS5B region. These strains, indicated with asterisks in Figure 5, appear to be 2k/1b recombinants and were both isolated from Georgian patients who stated as routes of infection intravenous drug use and sexual transmission. The first identified 2k/1b recombinant found was recovered in St. Petersburg in intravenous drug users (Kalinina *et al.*, 2002) and has since been found only in Estonia, Ireland, Uzbekistan and France in intravenous drug users (Kurbanov *et al.*, 2008a; Moreau *et al.*, 2006; Morel *et al.*, 2010; Tallo *et al.*, 2007). The putative recombinants found here cluster with each other and the St. Petersburg strain in both the Core-E1 and NS5B trees. Determination of recombination requires a more detailed analysis, as described in the second part of this study.

Analysis of both genomic regions revealed 16 strains for genotype 3, all of which belonged to subtype 3a, clustering with the 3a reference strains with bootstrap values higher than 95 in both trees. These strains were isolated from four Cypriot patients, four Russian patients, two Georgian patients, two Greeks, and one patient each from Bulgaria, Belarus and Pakistan. Because of the established correlation of genotype 3a with intravenous drug use (Morice *et al.*, 2006), these strains were further investigated in a new dataset context by comparing them to genotype 3a sequences from intravenous drug users available in the database. A new tree was constructed using the NS5B region of only the 3a strains from the general population group in this study (2005-2008) and 50 3a sequences derived from

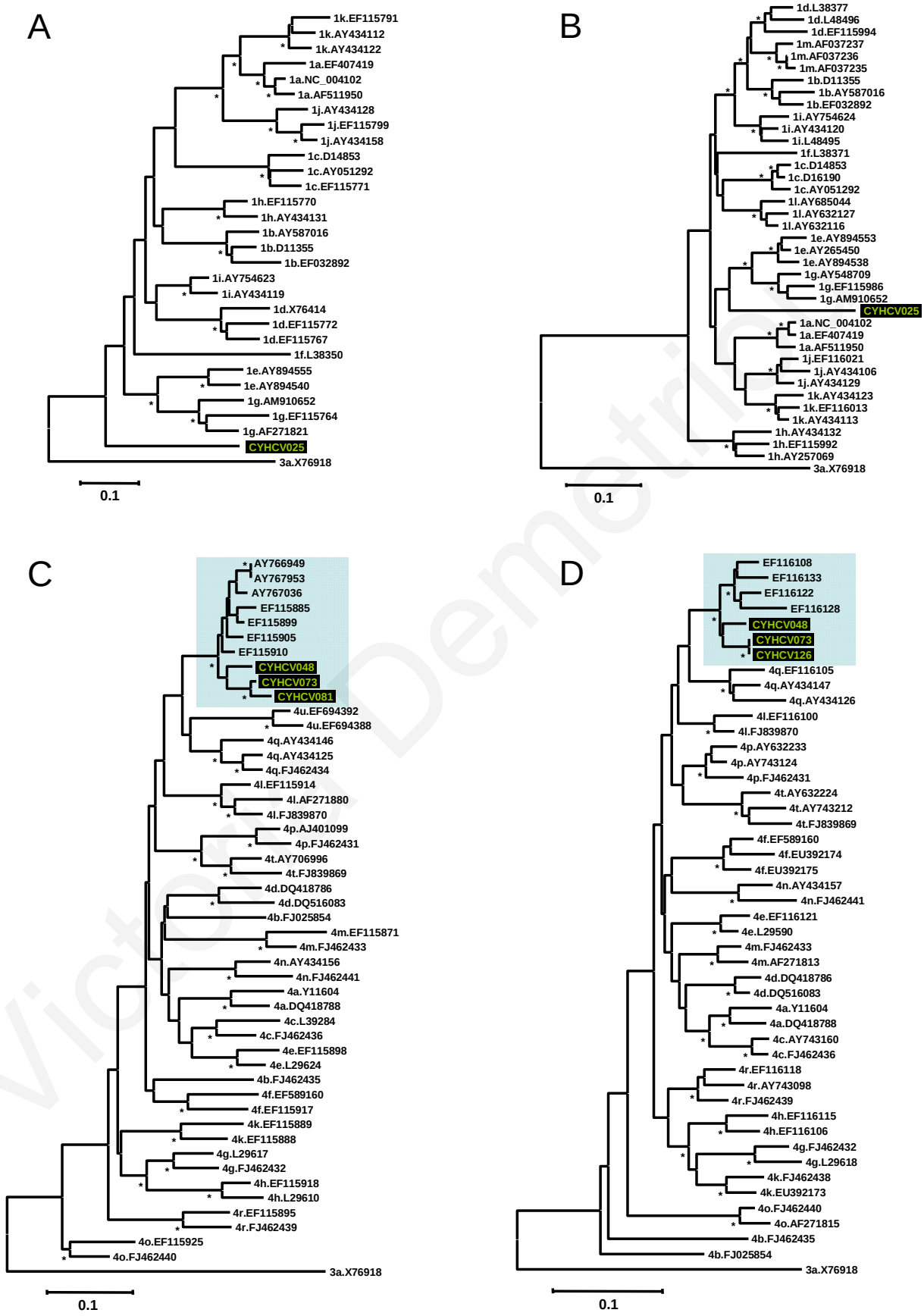

Figure 6 (previous page). Neighbour-joining trees showing the relationship of the unclassified strains with reference strains from all available subtypes of their genotype. (A) and (B) indicate the unclassified genotype 1 strain, CYHCV025, in the Core-E1 and NS5B regions, respectively. (C) and (D) indicate the unclassified genotype 4 strains CYHCV048, CYHCV073, CYHCV081 and CYHCV126, in the Core-E1 and NS5B regions, respectively. All trees are rooted with a genotype 3a strain. Asterisks at branch nodes signify percentage bootstrap support for 1,000 replicates that are above 70. The divergence between any two strains is estimated by the summation of the branch length using the scale at the lower left side of each tree.

publications (Cochrane *et al.*, 2002; Kalinina *et al.*, 2001; Morice *et al.*, 2006) and from intravenous drug users search in the HCV sequence database (Figure 7). This analysis yielded no significant evidence for the Cypriot strains clustering with each other or with isolates from intravenous drug users from other countries. As found previously (Morice *et al.*, 2006), the geographical areas of origin of the subjects do not seem to lead to distinct clusters of strains; it was therefore not possible to verify that the 3a strains found in this study were introduced by intravenous drug use in specific geographical regions. The observation of no country-specific phylogenetic clustering for strains isolated from intravenous drug users has been made for all genotypes (van Asten *et al.*, 2004). Also it has been shown previously that there is phylogenetic mixing of HCV subtype 3a strains from drug users and non-drug users in various countries, supporting the existence of a unique origin for subtype 3a (Cochrane *et al.*, 2002; Kalinina *et al.*, 2001; Morice *et al.*, 2006; Samimi-Rad *et al.*, 2004).

For genotype 4, six strains found in this study are assigned as subtype 4a in both regions. These strains were isolated from four Cypriot patients and two patients of Egyptian origin, which is the geographical area with a significantly high prevalence of HCV-4a (Abdel-Hamid *et al.*, 2007; Genovese *et al.*, 2005; Pybus *et al.*, 2003; Ray *et al.*, 2000; Tanaka *et al.*, 2004). It is significant to observe the origin of the patients with 4a infection and note the fact that this subtype was not isolated from patients of any other ethnic origin in this study group, which includes a high percentage of patients from countries of the former Soviet Union and other countries. This highlights the fact that this subtype is more restricted to a specific geographical radius than genotypes 1, 2, and 3.

Of particular interest is a cluster of a further four strains, isolated from two Cypriot patients, one Greek, and one Bulgarian patient, which was found to belong to genotype 4 but did not cluster with the standard reference sequences. Two of these samples were positive in both genomic regions, one was PCR positive only in the Core-E1 region, and the other only in the NS5B region. Consequently, a more detailed genotype-4-specific phylogenetic analysis was performed using reference sequences covering all available genotype 4 subtypes that have

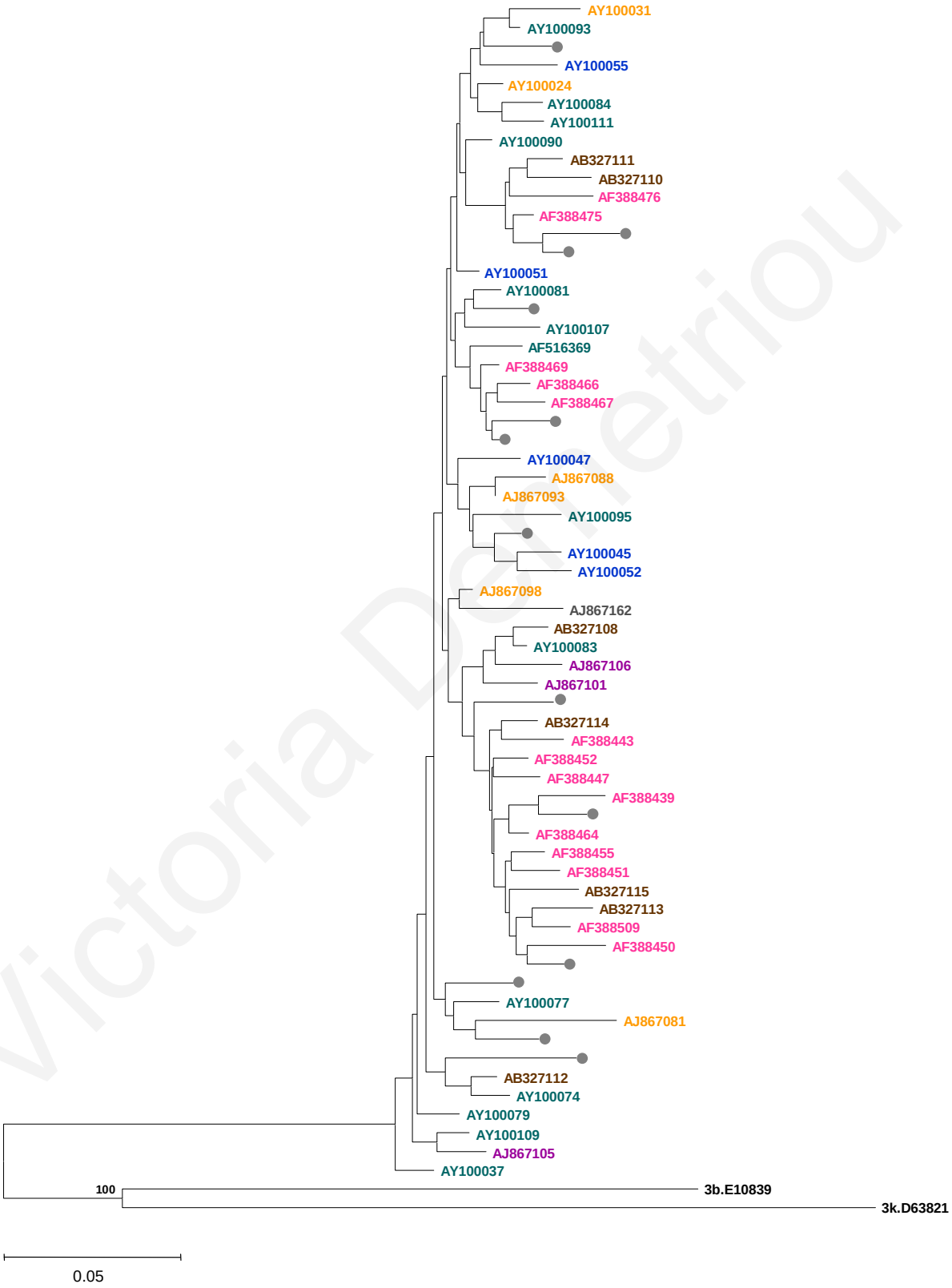

Figure 7 (previous page). Neighbour-joining phylogenetic trees for the partial NS5B sequences (positions 8316-8555, according to strain H77, Acc. No. NC_004102) of HCV subtype 3a strains obtained from patients within the general population of Cyprus. Trees were constructed using 50 intravenous drug use related subtype 3a sequences taken from published studies (Cochrane *et al.*, 2002; Kalinina *et al.*, 2001; Morice *et al.*, 2006) and from searches in the HCV sequence database of Los Alamos National Laboratory. The reference sequences are identified with their GenBank accession number and colour-coded, with the colours corresponding to the sampling country: Australia (yellow), France (blue), Russia (pink), Uzbekistan (brown), Brazil (gray), United Kingdom (dark green) and USA (purple). The Cypriot strains are indicated with gray circles. The numbers indicated at the nodes are percentage bootstrap support for 1,000 replicates. The divergence between any two sequences is obtained by summing the branch length, using the scale at the lower left of the tree.

sequences in these genomic regions (Figure 6C & D). The sequences from both regions were also uploaded into the HCV BLAST tool of the Los Alamos HCV sequence database and any strains with a sequence similarity of over 90% were also included in the phylogenetic tree construction. The strains in question do not group within any available assigned subtype of genotype 4 and were therefore labelled unclassified. This unclassified cluster appears to be genetically closest to subtypes 4q, 4m, and 4l. In the Core-E1 region, the unassigned strains from Cyprus cluster with seven other sequences from the database, four from subjects of African origin from a Canadian submission (Acc. No. EF115885, EF115899, EF115905, and EF115910) (Murphy *et al.*, 2007b) and three from patients of unknown epidemiological and demographic details from a UK submission (Acc. No. AY766949, AY767036, and AY767953), none of which are assigned a subtype in the database. No risk factors are kown for the additional sequences found in the database. A similar pattern is revealed in the NS5B tree, where the two sequences from this study cluster with the same four Canadian strains (Acc. No. EF116108, EF116122, EF116128, and EF116133), separately from any assigned strains. The British strains were not sequenced in the NS5B region, and as a result were not available for the NS5B analysis. This group of strains may belong to a novel subtype within genotype 4 and they are investigated in more detail in the second major section of the thesis.

Within genotype 5, just one strain was identified and this was classified as subtype 5a on both trees with corresponding bootstrap values of 100 in both cases. This is a subtype found primarily in South Africa (Chamberlain *et al.*, 1997b), but has also been found in west Flanders, Belgium (Verbeeck *et al.*, 2006), central France (Henquell *et al.*, 2004), and Syria (Antaki *et al.*, 2009). This strain is from a Cypriot patient who had a transfusion with 17 units (17x450 ml) of blood in 1975 in Johannesburg after a serious accident, but was only diagnosed with hepatitis C in 2006. From this data, it appears that this is a single 5a infection in Cyprus and, like the 2c strain, does not represent a spread of this genotype on the island.

Lastly, no strains of genotype 6 or 7 were identified in this dataset.

Overall, in the samples taken from HCV-infected subjects from the general population, a broad genotype distribution is revealed, and the phylogenetic analysis shows a high genetic diversity among the circulating strains. Strains of most genotypes were found, and subtype 1b was the most prevalent, followed by 3a, 1a, and 4a, and an unclassified type 4, with isolated cases of 2c, 5a and unclassified type 1 infections. It is worth noting that all genotypes and subtypes found in the general population group were found in patients of Cypriot nationality, except the putative 2k/1b recombinant, which was isolated from two Georgian subjects. This indicates that the high genetic heterogeneity is within the Cypriot population, and not just in foreign nationals currently residing on the island. In the patients coming from countries of the former Soviet Union, the HCV strains identified belonged to subtypes 1b, 1a, 3a, and the putative 2k/1b recombinant strain, which are representative of the HCV subtypes circulating in Eastern Europe (Kalinina *et al.*, 2001; Kurbanov *et al.*, 2003; Naoumov, 1999; Tallo *et al.*, 2007). In the Egyptian patients, only subtype 4a was found, reflecting the HCV situation in their country (Abdel-Hamid *et al.*, 2007). For patients from other European countries (Greece, UK, Italy, Romania, and Bulgaria) the genotypes identified were 1a, 1b, 3a, and 4, which are genotypes found commonly in western and southern Europe (Ansaldi *et al.*, 2005; Esteban *et al.*, 2008; Katsoulidou *et al.*, 2006; Trepo and Pradat, 1999). Finally, among patients of Asian ethnicity (Iranian and Pakistani), the genotypes discovered were 1a and 3a, respectively, again corresponding to the most prevalent types of HCV in their countries (Idrees, 2008; Samimi-Rad *et al.*, 2004). It is, however, difficult to make any further epidemiological conclusions, as for many patients the mode of transmission is unknown and, equally important, so is the country of infection.

The genetic diversity of HCV among the general population in Cyprus, as shown in this section of the study, is similar to the findings of HCV diversity in Greece (Katsoulidou *et al.*, 2006), and unlike the findings in Turkey, where subtype 1b is predominant (Altindis *et al.*, 2006; Altuglu *et al.*, 2008), Egypt, which has mainly subtype 4a (Abdel-Hamid *et al.*, 2007), or other countries in the Middle East, where genotypes 4 and 1 predominate (Ramia and Eid-Fares, 2006; Watson *et al.*, 1999). The heterogeneity in HCV genotypes found in Cyprus is probably due to imported strains from repatriated Cypriots, Cypriots travelling abroad and the large tourism industry. The diverse ethnic background of the study group and the finding of the possible 2k/1b recombinant strains also emphasise the impact of immigration from Eastern

Europe and increasing intravenous drug use in Cyprus on the multiple points of introduction and risk of widespread transmission of HCV strains on the island.

2.3.3.2. Intravenous drug users (2008)

Following the findings of subtypes globally associated with injecting drug use (3a, 1a and 2k/1b) within the general population in Cyprus (section 2.3.3.1), and knowing the high risk of HCV infection within this population, the investigation naturally proceeded to specifically sample intravenous drug users approached at rehabilitation centres all over the island. From the PCR-positive samples, sequences in both regions were obtained and included in phylogenetic analysis as described above. The strains from this cohort are seen in the trees of Figure 5, where they are indicated with blue triangles.

From studies carried out on intravenous drug users, the main genotypes found to be associated with this transmission route are 1a and 3a (Esteban *et al.*, 2008). The results of the subtyping of injecting drug user samples from Cyprus showed that the strains of this study group fell into only two subtypes. Eight samples (57%) belong to subtype 3a, and six samples (43%) belong to subtype 1b, with bootstrap values higher than 98. Two of the 3a strains are not included in the NS5B analysis, as they were PCR negative in this region. There was concurrence for all other samples between the two regions. Thus, the frequent presence of 3a strains in this population has been confirmed, but subtype 1a, which appears in the general population, was not found in any individuals of this cohort. Interestingly, 43% of the strains were subtype 1b, showing that this subtype has a significant prevalence in intravenous drug users on the island, and possibly illustrates spill-over between different risk groups in this geographical region and the countries where the strains have possibly originated from, that is, eastern and southern Europe. This would have to be confirmed by phylogenetic analysis with HCV sequences from these countries. About 10% of the sequences from intravenous drug users submitted into the Los Alamos HCV database (http://hcv.lanl.gov) are subtype 1b, which has been discovered in other such studies also, mostly at a lower prevalence though (Mathei *et al.*, 2005; Oliveira *et al.*, 2009; Peng *et al.*, 2008; Tallo *et al.*, 2007; Zhou *et al.*, 2006). Finally, no 2k/1b strains were found in this cohort, even though such isolates have been identified on the island and they are known to circulate among intravenous drug users (Kalinina *et al.*, 2002; Kurbanov *et al.*, 2008a; Kurbanov *et al.*, 2008b).

The phylogenetic analysis of the sequences resulted in no significant relationship between the injecting drug users' strains and those of the general population. Even with such a

permissive arbitrary threshold of genetic distance as 0.05, there were no clusters observed, confirming the polyphyletic nature of HCV infection on the island and illustrating that the HCV epidemic in the selected population runs in parallel to that in the general population infected with HCV.

Three small clusters were observed within the intravenous drug users group itself. However, with small genetic distances (< 0.05) and high bootstrap values (Figure 5). These clusters are between samples CYIDU06, CYIDU07, and CYIDU33 within genotype 1b, CYIDU24 and CYIDU37 also within 1b, and CYIDU20, CYIDU26, and CYIDU27 within genotype 3a. In the first cluster (CYIDU06, CYIDU07, and CYIDU33), the sequences were obtained from three males between the ages of 24 and 47, all sampled from the same town, two of which from the same centre. Two were Cypriot and the other was Greek. They all stated that they had practiced intravenous drug abuse within the past year, and interestingly all had been practicing intravenous drug use for over 10 years. In the second cluster (CYIDU24 and CYIDU37) the samples were taken from a 27-year-old Bulgarian male and a 31-year-old Cypriot male, sampled in different towns. They both mention speedball as the drug used when first injected drugs. Both have been practicing intravenous drug users for over 13 years. Only one other individual reported this, CYIDU02, however this person tested negative in the HCV antibody. The third cluster (CYIDU20, CYIDU26, and CYIDU27) includes samples from three Cypriots, one female and two males aged between 19 and 24. Two of the samples were obtained from users in the same detoxification clinic and the other from another centre in another town. The subjects from the same detoxification clinic stated a duration of intravenous drug abuse of less than five years and the other individual did not state how long he had been using. These small clusters indicate the probability of injecting equipment being shared between these users. Also, Cyprus is a small island and easy to get around, with a relatively small injecting drug user community. Therefore it is not surprising that strains cluster from users who were sampled in different towns or rehabilitation centres.

The Bayesian analysis of the Cypriot intravenous drug users' sequences with the most genetically related database sequences, retrieved from the BLAST search, produced no significant clustering, as indicated by posterior probabilities that did not exceed 0.90. Lack of sequence data in the database impairs the ability to investigate molecular epidemiological and phylogenetic relationships between newly obtained sequences and HCV sequences from anywhere around the world. From these results it cannot be determined where the HCV infections originated from or how they made their way to Cyprus through intravenous drug

users. More sequence data from intravenous drug users, especially from eastern and southern European countries, would be required to obtain a more accurate picture of the HCV epidemic and to gain insight into the origin and dynamics of HCV infection at the population level. This point is the most severe limitation to this study and to all work regarding phylogenetic analysis of HCV strains.

2.3.3.3. Incarcerated population (2009-2010)

From 25 samples taken for diagnostic testing from individuals serving a prison sentence at the state prisons, 13 were PCR-positive in the Core-E1 region and 16 in the NS5B region. The sequences were included in the phylogenetic analysis of strains recovered in Cyprus and are seen as pink squares in the trees in Figure 5. All sequences were clearly grouped within subtypes and the genotype distribution was as follows: 9 were 3a (56%), 5 were 1b (31%), and 2 were 1a (13%). This shows a majority agreement with the genotype distribution in the Cypriot intravenous drug user population (section 2.3.3.2), except for the presence of 1a strains, which are commonly associated with injecting drug abuse.

Certain small phylogenetic clusters (genetic distances < 5%) were observed between strains from this cohort and those of the intravenous drug users. Within subtype 3a, sample HCVP05, from a 23 year old Cypriot male, clusters within an existing cluster of strains from Cypriot intravenous drug users (two male, one female) in both trees. The subjects of this cluster are also of similar age (19-24 years). Within the same subtype, two separate clusters are seen between strains of the prisoners' cohort. One is formed by samples HCVP04, HCVP07 and HCVP09, derived from one female and two male subjects, aged 21-35 years, two Cypriot and one of unknown nationality. This cluster is only observed in the NS5B tree, as two of the samples were PCR-negative in the Core-E1 region. The other small cluster is formed in both trees by samples HCVP15 and HCVP16, from two male subjects ages 39 and 49, one Cypriot and one of unknown nationality.

Within subtype 1b, sample HCVP17, from a 42 year old Cypriot male, groups closely with an existing cluster of strains from Cypriot and Greek male injecting drug users in the trees of both genomic regions. It is interesting that the Cypriot subjects who cluster together here are in a similar age group (39-47 years) that indicates the probability of a long duration of drug use.

In the same subtype, another small cluster is observed, involving sample HCVP12, from a 24 year old Georgian male, and a strain in the general population, CYHCV93, from a

39 year old Georgian male who was infected through intravenous drug use and, interestingly is a 2k/1b recombinant strain, known to circulate in certain injecting drug user circles in Eastern Europe (Kurbanov *et al.*, 2008b). However, sample HCVP12 was PCR negative in the Core-E1 region, and so it cannot be ascertained if it was a 2k strain in this region and whether it would cluster with a similar sample from the general population. This phylogenetic relationship in the NS5B region, and the similar demographic and epidemiological characteristics (Georgian nationality, male, 25 years old, history of intravenous drug use), could mean that the strain in question is also a recombinant of this type, but this would have to be confirmed with additional sequence data.

Overall, epidemiologically, the observations indicate that users practise in small groups and these inmates have had contact with the intravenous drug users discussed above, either directly or indirectly. It can be seen that these small clusters are formed between strains from individuals who have common characteristics (nationality, age group) and injecting drug use is the common route of infection for all, demonstrating that users with common nationality or within certain age groups are more likely to associate with each other, increasing the danger of sharing injecting drug use equipment. The results from these data highlight the significant issue of uncontrolled transmission between these injecting drug users and inmates.

2.3.3.4. HIV/HCV coinfected cohort (2005-2009)

From samples collected at the Laboratory of Biotechnology and Molecular Virology for an ongoing study of the molecular epidemiology of HIV on the island, fourteen had tested positive for HCV antibodies as well as HIV, and were subsequently included in this thesis to investigate the HCV sequence diversity among this group. Eight of these samples were PCR positive for both regions, and sequences could be obtained from seven samples. The sequences were included in the overall phylogenetic analysis and they are indicated with a green rhombus symbol in the trees of Figure 5. The subtypes found in this group of subjects were 1a (3 samples), 1b (2 samples) and 3a (2 samples). Two strains cluster closely together (genetic distance < 1%) in subtype 1b in both trees, and they are derived from two individuals who were a heterosexual couple at the time of sampling. Their HIV-1 strains also display a similar clustering both in the full genome analysis (Kousiappa *et al.*, 2009b) and in investigations of the gag, pol (Pr/RT) and env regions separately (Kousiappa *et al.*, 2009a). This is a strong indication of transmission from one individual to the other, but it cannot be certain whether infection was through sexual practises or household transmission. No other strains from this

cohort form any clusters and none show significantly close phylogenetic relationships with samples from neither the general population nor the other high-risk cohorts. Regarding the country of origin, the Cypriot study subjects in this cohort carried HCV genotypes 3a and 1a, Georgians carried 3a and 1b, and both the Latvian and Bulgarian subjects were infected with HCV subtype 1a. Concerning risk behaviours, the MSM carried HCV genotypes 3a and 1a. The subjects who were infected with HIV through heterosexual contacts were infected with all three HCV subtypes found in this cohort, possibly signifying the fact that they may have acquired HCV through different routes, or simply representing the fact that there were more individuals in this subgroup than for other HIV transmission routes in this cohort. The intravenous drug user whose HCV strain was sequenced was infected with 1a, a strain typically found in injecting drug users. Other studies have demonstrated the circulation of genotypes 1, 3 and 4 in both MSM and injecting drug users who are coinfected with HIV (Ghosn *et al.*, 2008; Liu *et al.*, 2008; Ponamgi *et al.*, 2009; Pukhrambam *et al.*, 2007; Re *et al.*, 2008; Urbanus *et al.*, 2009; van Asten *et al.*, 2004), however genotype 4 was not found in this cohort. Due to the small sampling number in this cohort, it is impossible to make any statistically significant epidemiological conclusions regarding association of genotype with possible route of infection.

3. Part II: Near-full genome sequencing and characterisation of selected isolates

3.1. Specific objectives

In the previous section, certain uncommon isolates were uncovered among the HCV-infected population in Cyprus. These strains include two putative 2k/1b recombinants, one unclassified genotype 1 isolate and a cluster of genotype 4 strains that did not group within any known subtype. Following these observations, the aim in this section was to develop strain-specific protocols for the RT-PCR and sequencing of the near-full genome of these isolates, in order to genetically analyse them in more detail, determine their classification status, and to submit them to GenBank, so that they would be publicly available in the HCV sequence databases to contribute to the global data on HCV sequence diversity and classification.

3.2. Methods

3.2.1. Samples

During the study of the molecular epidemiology of HCV infection in the general population of Cyprus, blood samples from HCV seropositive individuals were collected and two distinct regions at the two extremes of the HCV genome (Core-E1 and NS5B) were sequenced and analysed in all PCR-positive samples. The subtype classification of all samples which were PCR-positive for both regions was in agreement, except for two samples, which were found to be subtype 2k in the Core-E1 region, and 1b in the NS5B region (indicated with asterisks in Figure 5). These strains were considered putative 2k/1b recombinants, but further analysis was required to confirm this. Also in the general population group, one strain was found to belong in the genotype 1 group, but did not demonstrate phylogenetic association with any known subtypes of this group in either genomic region (Figure 6A & B). The third type of strain was also identified in the general population cohort and belongs to genotype 4. Two samples were sequenced in both regions, one in Core-E1 only, and one in NS5B only. These sequences formed clusters with each other, distinct from any known genotype 4 subtype with available sequences in the genomic regions investigated (Figure 6C & D). For this section of the study, stored plasma (-70^{o}C) from the previously collected samples was used.

3.2.2. Experimental design

A protocol and primers for RT nested PCR and sequencing of overlapping fragments spanning the near-full genome were designed based on a published experimental design (Noppornpanth *et al.*, 2006). For the 2k/1b strains, PCR primers were designed to be subtype-specific, by using alignments of available 2k/1b, 2k, and 1b sequences from the Los Alamos HCV sequence database (http://hcv.lanl.gov). With the assumption that the strains in question would have the same recombination point as the first 2k/1b recombinants found in St. Petersburg in 2002 (Kalinina *et al.*, 2002), where the crossover point was found to be within the NS2 region around nucleotide 3164 (position according to strain H77), primers on the 5′ side of this point were designed to be 2k-specific and primers on the 3′side were designed to be 1b-specific. Also, the region covering the putative crossover point was sequenced bidirectionally from three different PCR products to ensure agreement between sequences. For the unclassified genotype 1 strain, primers were designed from alignments of non-a, non-b subtypes of genotype 1 sequences available on the database, and for the unclassified genotype 4 isolates, primers were designed based on alignments of genotype 4 sequences from the database with more emphasis on the genetically closest subtypes (4l, 4p, 4q, 4t, 4u). In all cases, in the regions where sequences of the strains were available from the previous study (Core-E1 and NS5B, section 2.3.3.1), the primers were designed specifically using these sequences. For the resulting PCR products, the inner PCR primers were also used for initial sequencing of PCR products and further strain-specific sequencing primers were subsequently designed by "primer walking", using the newly derived sequences. The primers were designed to have similar thermodynamic properties (nucleotide length and GC content), so as to maintain a single annealing temperature in the thermal cycling conditions of all PCR and cycle sequencing reactions. The RT-PCR and sequencing designs for the two 2k/1b recombinants are seen schematically in Figure 8 and Figure 9, for the unclassified genotype 1 strain in Figure 10, and for the unclassified genotype 4 strains in Figure 11. All primers designed and used in this study for the near full genome amplification and sequencing of the 2k/1b recombinant, the unclassified genotype 1 and the unclassified genotype 4 strains are listed in Table 4, Table 5 and Table 6, respectively.

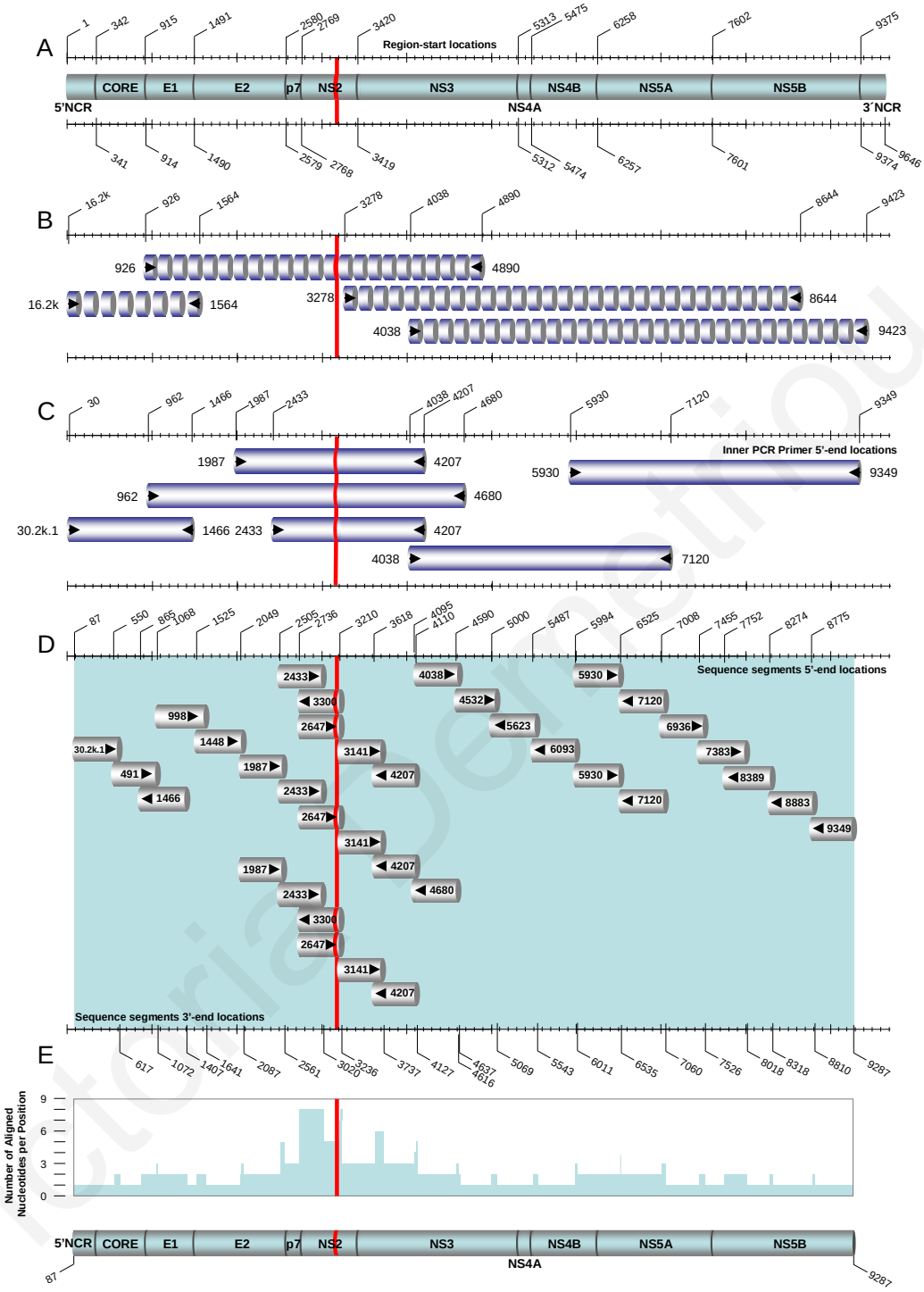

Figure 8. PCR and sequencing design for the 2k/1b recombinant strain CYHCV037. (A) HCV genome map numbered according to strain H77 (Acc. No. NC_004102); (B) the outer PCR products, the numbers on the scale above indicating the beginning and end of each, the numbers on either side of the cylinders indicating the names of the primers used; (C) the inner PCR products, the numbers on the scale above indicating the beginning and end of each product, the numbers on either side of the cylinders indicating the names of the primers used; (D) the overlapping segments read by sequencing are shown as cylinders, the name and arrow in each segment indicating the sequencing primer and directionality of the reading, the locations on the scale above and below indicating the beginning and end locations of the readings, respectively; (E) a scale designating the number of overlapping sequence readings for each position on the near-full genome, below which a genome map showing the total sequence retrieved. The vertical red line designates the location of the recombination point.

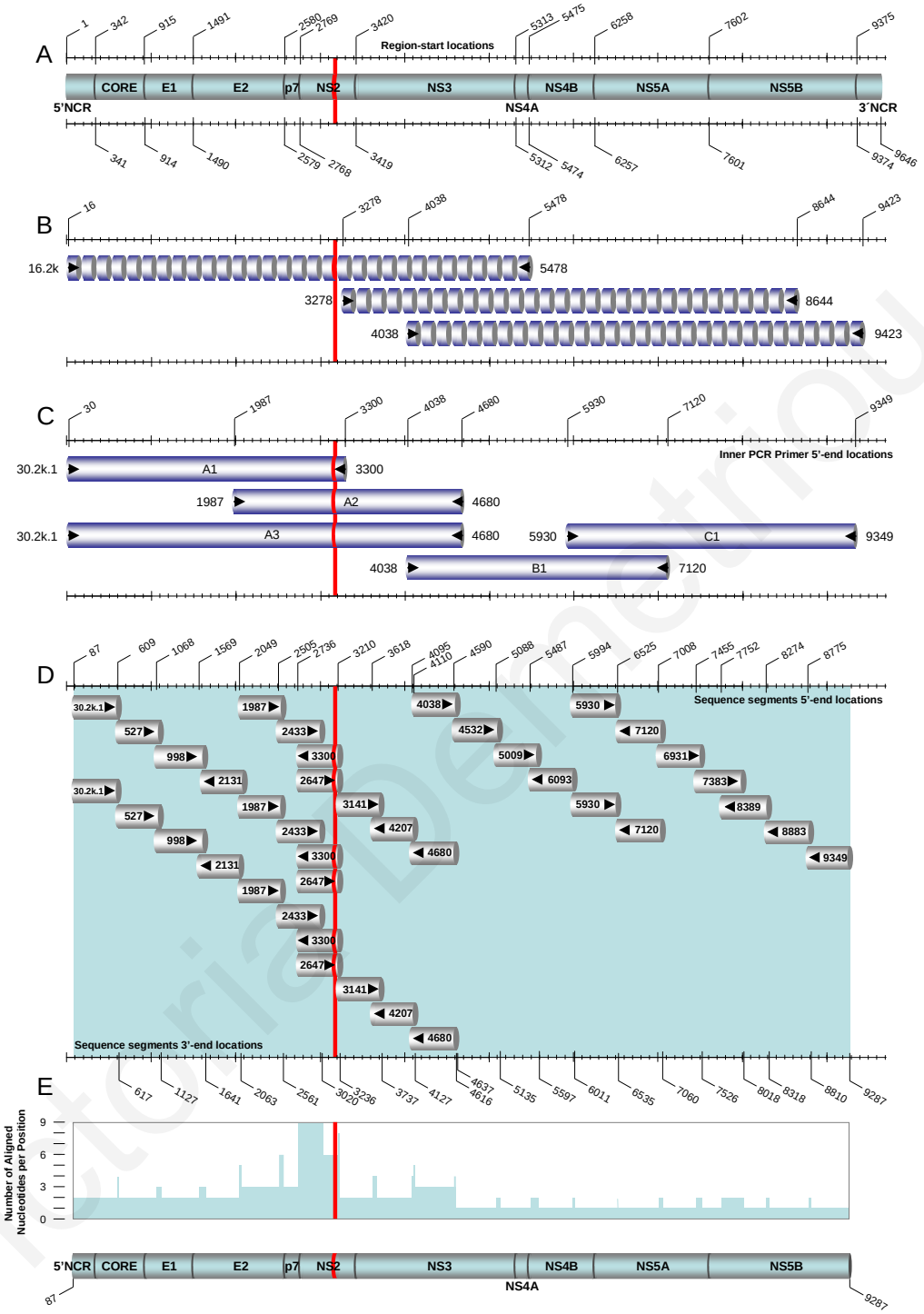

Figure 9. PCR and sequencing design for the 2k/1b recombinant strain CYHCV093. (A) HCV genome map numbered according to strain H77 (Acc. No. NC_004102); (B) the outer PCR products, the numbers on the scale above indicating the beginning and end of each, the numbers on either side of the cylinders indicating the names of the primers used; (C) the inner PCR products, the numbers on the scale above indicating the beginning and end of each product, the numbers on either side of the cylinders indicating the names of the primers used; (D) the overlapping segments read by sequencing are shown as cylinders, the name and arrow in each segment indicating the sequencing primer and directionality of the reading, the locations on the scale above and below indicating the beginning and end locations of the readings, respectively; (E) a scale designating the number of overlapping sequence readings for each position on the near-full genome, below which a genome map showing the total sequence retrieved. The vertical red line designates the location of the recombination point.

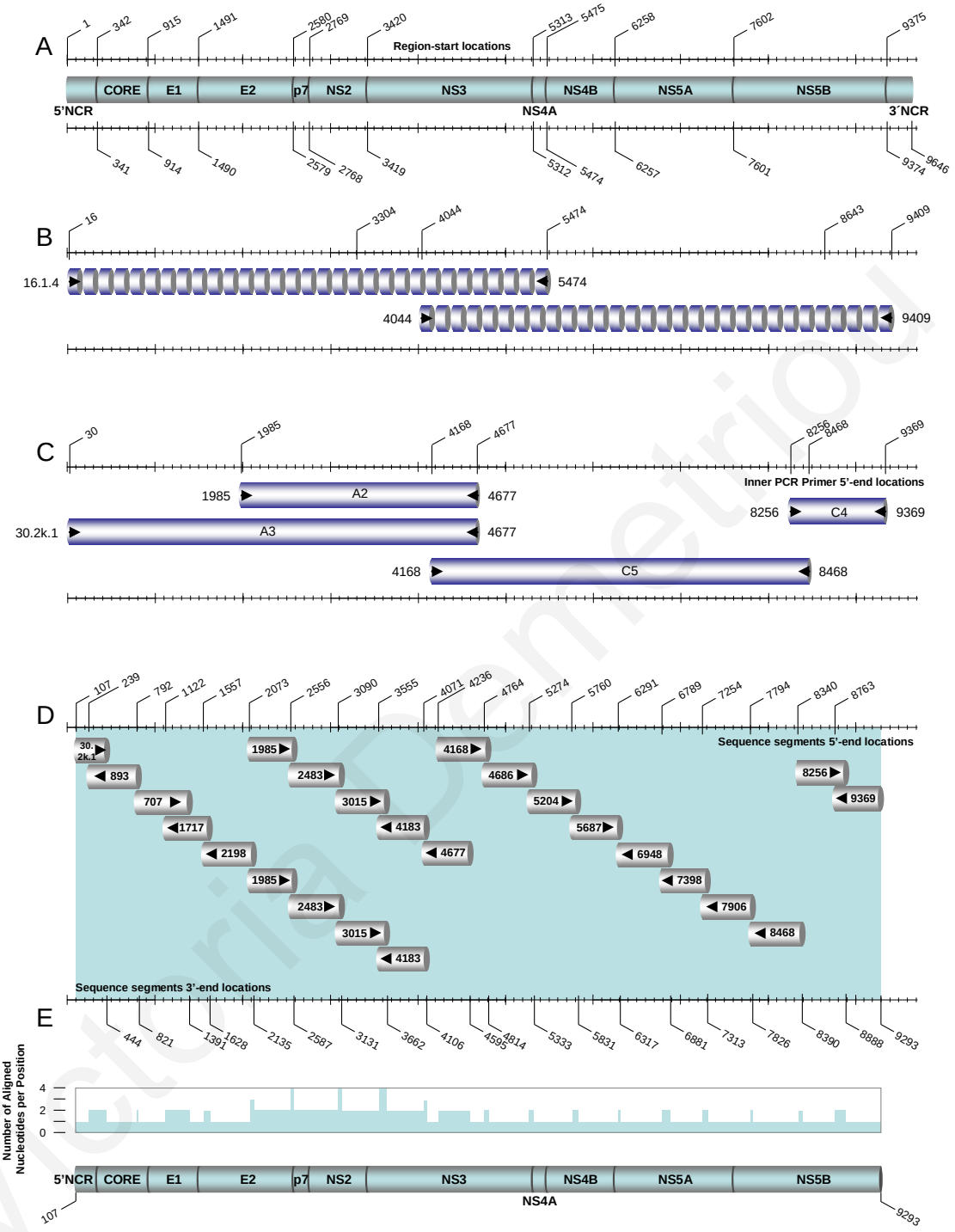

Figure 10. PCR and sequencing design for the unclassified genotype 1 strain. (A) HCV genome map numbered according to strain H77 (Acc. No. NC_004102); (B) the outer PCR products, the numbers on the scale above indicating the beginning and end of each, the numbers on either side of the cylinders indicating the names of the primers used; (C) the inner PCR products, the numbers on the scale above indicating the beginning and end of each product, the numbers on either side of the cylinders indicating the names of the primers used; (D) the overlapping segments read by sequencing are shown as cylinders, the name and arrow in each segment indicating the sequencing primer and directionality of the reading, the locations on the scale above and below indicating the beginning and end locations of the readings, respectively; (E) a scale designating the number of overlapping sequence readings for each position on the near-full genome, below which a genome map showing the total sequence retrieved.

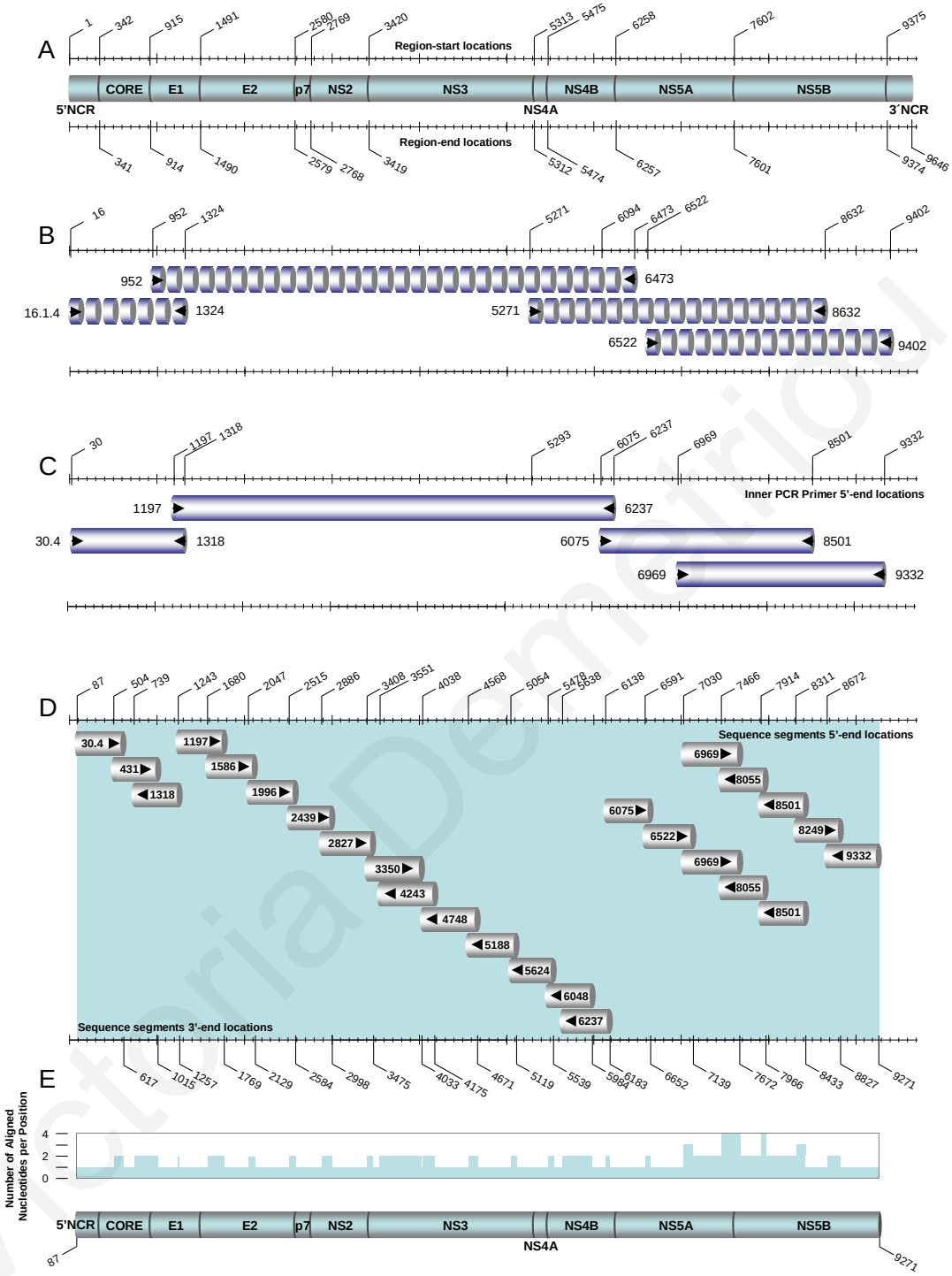

Figure 11. PCR and sequencing design for the unclassified genotype 4 strains. (A) HCV genome map numbered according to strain H77 (Acc. No. NC_004102); (B) the outer PCR products, the numbers on the scale above indicating the beginning and end of each, the numbers on either side of the cylinders indicating the names of the primers used; (C) the inner PCR products, the numbers on the scale above indicating the beginning and end of each product, the numbers on either side of the cylinders indicating the names of the primers used; (D) the overlapping segments read by sequencing are shown as cylinders, the name and arrow in each segment indicating the sequencing primer and directionality of the reading, the locations on the scale above and below indicating the beginning and end locations of the readings, respectively; (E) a scale designating the number of overlapping sequence readings for each position on the near-full genome, below which a genome map showing the total sequence retrieved.

3.2.3. Reverse transcription nested PCR and sequencing

For all samples, RNA was extracted from 200 µl plasma that was previously extracted from the blood obtained at the initial sampling from the patients (section 2.2.2) using the High Pure Viral RNA kit (Roche Diagnostics, Manheim, Germany), following the manufacturer's instructions. Initially, 50 µl one-step RT-PCR reactions were set up using the Superscript™ III one-step RT-PCR system with Platinum® Taq high fidelity (Invitrogen, Carlsbad, CA, USA) and using 15 µl RNA, 1X reaction mix, 1 µl of Superscript III™ RT/Platinum® Taq High Fidelity enzyme mix and 20 pmol each of the outer forward and reverse primers. Cycling conditions were as follows: a reverse transcription step at 56 ^{o}C for 60 min, and 94 ^{o}C for 2 min, followed by 43 cycles at 94 ^{o}C for 15 s, 56 ^{o}C for 30 s and 68 ^{o}C for 1 min per kilobase of product, and a final elongation step at 68 ^{o}C for 5 min. For the nested PCR reactions, the Platinum® PCR SuperMix High Fidelity (Invitrogen, Carlsbad, CA, USA) was used to make a 50 µl reaction with 3 µl of the corresponding first round PCR product, 1X reaction mix and 20 pmol each of the inner forward and reverse PCR primers. Cycling conditions were as follows: 94 ^{o}C for 2 min, and then 40 cycles at 94 ^{o}C for 20 s, 56 ^{o}C for 30 s and 68 ^{o}C for 1 min per kilobase of product, and a final elongation step at 68 ^{o}C for 5 min. PCR products were visualised after electrophoresis on a 2% agarose gel stained with ethidium bromide, and purified using the QIAquick ® PCR Purification Kit (Qiagen, Hilden, Germany), following the manufacturer's instructions. The concentration of each PCR product was measured at 260 nm on the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Cycle sequencing reactions were performed on 1µl of each secondary PCR product using the BigDye ® Terminator v3.1 cycle sequencing kit (Applied Biosystems, Warrington, UK) with 5 pmol of each sequencing primer, as described above (section 2.2.4), and subsequently cleaned using NucleoSEQ columns (Macherey-Nagel, Düren, Germany), according to the manufacturer's instructions. Direct population sequencing was carried out in order to obtain consensus sequences reflecting the dominant quasispecies population in each sample. Sequences were read using the ABI 3130 Genetic Analyzer and obtained through the ABI sequencing analysis 5.2 software (Applied Biosystems, Foster City, CA, USA). The results were manually checked and polymorphisms where double peaks were observed of equal intensity in the chromatogram were annotated with standard IUB base codes accordingly

Table 4. Primers for the near-full genome amplification and sequencing of 2k/1b recombinant strains.

Designation [a]	Sequence [b]	Position [c]
PCR primers		
16.2k (F)	GGGGCGACACTCCGC	16-30
30.2k.1 (F) [d]	CCATGAATCACTCCCCTGTGAGG	30-52
926 (F)	GAACACCAGTCAGACCTACATGG	926-948
962 (F)	CTCTAACAGCAGCATCACGTGG	962-983
1466 (R) [d]	GAGGATGACGACGACCTTGG	1447-1466
1564 (R)	GAAAAGGCTGGTAAGGCGAAGC	1543-1564
1987 (F) [d]	GGTTTGGATGCACATGGATGAATGG	1987-2011
2433 (F) [d]	GACGTGCAGTACTTGTATGGCC	2433-2454
3278 (F)	GGAGACCAAGATCATCACCTGGG	3278-3300
3300 (R) [d]	CCCAGGTGATGATCTTGGTCTCC	3278-3300
4038 (F) [d]	GGCAGCGGCAAGAGCAC	4038-4054
4207 (R) [d]	GCACCTGTGGTGATGGTCC	4189-4207
4680 (R) [d]	CGCCGGTRWAGCCCGTC	4664-4680
4890 (R)	GGACCGAGGAGTCAAACATGC	4870-4890
5478 (R)	AGGCGCACTCTTCCATYTCATC	5457-5478
5930 (F) [d]	CGTGGCCTTTAAGGTYATGAGCG	5930-5952
7120 (R) [d]	CGRAGCGGGTCRAAAGAGTC	7101-7120
8644 (R)	GCTGAGTACCTAGTCATAGCCTCCG	8620-8644
9349 (R) [d]	ACCCCTACAGAAAGTAGGAGTAGGC	9325-9349
9423 (R)	AAAAAACAGGATGGCCTATTGGCC	9397-9423
Sequencing Primers		
491 (F)	GAAGACTTCCGAGCGGTCC	491-509
527 (F)	CCAGCCCATCCCTAAAGATCG	527-547
998 (F)	GGTACTACATGTCCCTGGATGC	998-1019
1448 (R)	CAAGGTCGTCGTCATCCTCC	1448-1467
2131 (R)	CATTTGGTGTAAGTAGCCTCAGGG	2108-2131
2647 (F)	CCATCTTCTTCATAGCTGCCTGG	2647-2669
3141 (F)	GTCCAGATGATGCTGCTAACCC	3141-3162
4532 (F)	GAAGAAATGTGATGAGCTCGCTGC	4532-4555
5009 (F)	GTTCTGGGAGAGCGTCTTCAC	5009-5029
5623 (R)	CAGAAGGTCTCAAGGGCTCG	5604-5623
6093 (R)	GGTTCATCCACTGCACAGCC	6074-6093
6931 (F)	CCCTGGCTAGCTCTTCTGC	6931-6949
6936 (F)	GCCAGCTCTTCAGCTAGTCAG	6936-6956
7383 (F)	GCTACAAAGACCTTCGGCAGC	7383-7403
8389 (R)	CCGATATAAAGTCGCTCTGTGAGC	8366-8389
8883 (R)	GAGTCATCAGAATCATCCTTGCCC	8860-8883

[a] Orientation of the PCR primer is indicated in parenthesis: F, forward; R, reverse.
[b] IUB base codes annotate equimolar mixtures of bases in the oligonucleotides: Y, C and T; R, A and G; W, A and T.
[c] Positions are numbered according to HCV strain H77 (Acc. No. NC_004102).
[e] Also used as sequencing primers.

Table 5. Primers for near-full genome amplification and sequencing of the unclassified genotype 1 strain.

Designation [a]	Sequence [b]	Position [c]
PCR primers		
16.1.4 (F)	GGGGCGACACTCCACC	16-31
30.2k.1 (F) [d]	CCATGAATCACTCCCCTGTGAGG	30-52
1985 (F) [d]	TTGGTTCGGTTGYACCTGGATG	1985-2006
4044 (F)	GGCAAGAGCACSAAGGTGC	4044-4062
4168 (F) [d]	CAAATATCAGGACCGGTGTCAGG	4168-4190
4677 (R) [d]	CGGTGWAGCCAGTCATBARGGC	4656-4677
5474 (R)	GCACTCYTCCATCTCATCRAACTC	5451-5474
8256 (F)	TATGAYACCCGCTGYTTTGACTCNAC	8256-8281
8468 (R)	GAGGCTGGTCGTTAGGACAC	8449-8468
9369 (R) [d]	TGGGRAGSAGRTAGATGCCTACC	9347-9369
9409 (R)	GGCCTAAGAGGCCTGGAGTG	9391-9409
Sequencing Primers		
707 (F)	CATCGATACCCTCACGTGTGG	707-727
893 (R)	GCAAGATAGTAAGGCCAGAAGGAAG	869-893
1717 (R)	CAGCTGCTCATCCTCTCAGG	1698-1717
2198 (R)	CGGATAGTGCCACAGCCTATAAG	2176-2198
2483 (F)	CAAGTGGGAGTACGTGGTGC	2483-2502
3015 (F)	CTGCTGGCTATTCTGGGACC	3015-3034
4183 (R)	CCGGTCCTGATATTTGGATCAAGC	4160-4183
4686 (F)	GACTCGGTGATAGATTGCAACACC	4686-4709
5204 (F)	AACGCCTCTCCTATATAGGCTGG	5204-5226
5687 (F)	TGGTAATCCAGCTGTTGCATCG	5687-5708
6948 (R)	CAGAAGAACTGGCCAGAGAGG	6928-6948
7398 (R)	CAAATGTCTTCTTGGCGAGGTCG	7376-7398
7906 (R)	CCATATTTGGATCTGGCTGAGTGG	7883-7906

[a] Orientation of the PCR primer is indicated in parentheses: F, forward; R, reverse.
[b] IUB base codes annotate equimolar mixtures of bases in the oligonucleotides: R of A and G; S of C and G; W of A and T; B of C, G and T ;N of all bases.
[c] Numbering according to HCV H77 strain (Acc. No. NC_004102).
[d] Also used as sequencing primers.

in the resulting sequence. All resulting sequences from each strain were used for stitching together the near-full genome, as seen in part E of Figure 8, Figure 9, Figure 10 and Figure 11.

3.2.4. Near-full genome sequence analysis

Analysis and characterisation of the strains was done using the near-full genome sequence obtained from the RT-PCR and direct sequencing as described above. Each type of strain was analysed separately.

Table 6. Primers for near-full genome amplification and sequencing of unclassified genotype 4 strains.

Designation [a]	Sequence [b]	Position [c]
PCR primers		
16.1.4 (F)	GGGGCGACACTCCACC	16-31
30.4 (F) [d]	CCATRRATCACTCCCCTGTGAGG	30-52
952 (F) [d]	CAAATGACTGCCCAAACTCRAGC	952-974
1197 (F)	CTAGTGGGCCAACTGTTCACC	1197-1217
1318 (R) [d]	CAGTTCATCATCATRTCCCAWGCCATNCGRTGDCC	1284-1318
1324 (R) [d]	GGBGACCARTTYAKCATCATRTCCCAWGCC	1295-1324
5271 (F)	TACATCATGGCHTGCATGTCNGC	5271-5293
6075 (F)	GCNGTGCAGTGGATGAACCG	6075-6094
6237 (R)	CGTTGATCCACTTGTGGAGACG	6216-6237
6473 (R) [d]	CATCGAGCCGTTCTTGATGTGG	6452-6473
6522 (F)	CCCATCAATGCTTACACCACAGG	6522-6544
6969 (F)	TCCCTAAARGCCACATGCACC	6969-6989
8501 (R)	CGTGGCCTTAAGGTAGCACG	8482-8501
8632 (R)	GTCATAGCCTCCGTGAAGGC	8613-8632
9332 (R)	GAGTAGGCACAGGAGTAAATAGCGG	9308-9332
9402 (R)	ARGGTCGGAGTGTTAAGCTGC	9382-9402
Sequencing Primers		
431 (F)	CGTTGGCGGAGTTTACTTGTTGC	431-453
1586 (F)	CACCAACGGAAGCTGGCAC	1586-1604
1996 (F)	GTGTTTGGATGAACGGCACTGG	1996-2017
2439 (F)	CAGTACTTGTATGGTGTGAGCTCG	2439-2462
2827 (F)	TGACGCTATCCCCATACTACAAGC	2827-2850
3350 (F)	GGCTAGGCTGGGCAATGAG	3350-3368
4243 (R)	GCCARCAATTTCCCGTAGGTGG	4222-4243
4748 (R)	GGTTGGATCCAGGCTGAAATCC	4727-4748
5188 (R)	GGCTTGAGGCGGATCAGG	5171-5188
5624 (R)	CCAGAACTGCTCGAGTTTCTCG	5603-5624
6048 (R)	GCAGRATTGCTGCGCACAC	6030-6048
8055 (R)	CCTTCTCCGGTTTGACGGAG	8036-8055
8249 (F)	GTTCTCRTATGACACCCGCTGC	8249-8270

[a] Orientation of the PCR primer is indicated in parentheses: F, forward; R, reverse.
[b] IUB base codes annotate equimolar mixtures of bases in the oligonucleotides: Y of C and T; R of A and G; H of A, C and T; B of C, G and T; K of G and T; W of A and T; D of A, G and T; N of all bases.
[c] Numbered according to HCV H77 strain (Acc. No. NC_004102).
[d] Also used as sequencing primers.

Simplot and Bootscan analyses were carried out to visualise the percentage similarity of each strain with reference sequences along the length of its sequence, and to confirm or rule out recombination events. The near-full genome sequences were aligned along with reference strains representing all HCV genotypes and all available subtypes of the appropriate genotypes in each case in MEGA v4 (Tamura *et al.*, 2007) and uploaded into SimPlot v3.5.1 software

(http://sray.med.som.jhmi.edu/SCRoftware). In the case of the 2k/1b strains, the analysis also included the two available full genome sequences of 2k/1b strains, accession numbers AY587845 (Kalinina *et al.*, 2004) and FJ821465 (Morel *et al.*, 2010). The analyses were carried out with windows of 200 bases proceeding in steps of 20 bases, using the Kimura 2-parameter distance model, and a Neighbour-Joining tree model with 100 bootstrap replicates.

Following the results of the SimPlot and Bootscan graphs, phylogenetic analyses were also carried out. In the case of the 2k/1b strains, phylogenetic analyses were carried out separately on the genomic regions on either side of the crossover point (positions 87-3150 and 3200-9287, respectively, numbered according to strain H77), in order to confirm the classification of each region in the corresponding subtypes, by including reference strains and the most closely related (>92% similarity) sequences derived from a HCVBLAST search (http://hcv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). The analysis was also carried out on the whole derived sequence using the available 2k/1b genomes from GenBank and reference strains from the parental subtypes.

For the unclassified genotype 1 and 4 strains, the trees were constructed using reference sequences of all genotypes and all available subtypes of the corresponding genotypes. A HCV-BLAST search retrieved no near-full genomes with >76% or >84% sequence similarity, respectively. Sequences were uploaded in FASTA format and aligned in MEGA v4 (Tamura *et al.*, 2007) where a Neighbour-Joining tree (Saitou and Nei, 1987) was constructed using the Kimura 2-parameter distance estimation approach (Kimura, 1980) and reliability of the phylogenetic clustering was evaluated using 1,000 bootstrap replicates (Felsenstein, 1985).

3.2.5. Reference sequences

The GenBank accession numbers of the reference strains used in the construction of phylogenetic trees described above, are seen in the corresponding figures in the results section (section 3.3). For the Simplot and Bootscan analyses, the GenBank accession numbers of the general reference strains used were: 1a: NC_004102, EF407419, AF511950; 1b: AY587016, D11355, EF032892; 1c: D14853, AY051292; 2a.AB047639, AY746460, D00944; 2b.AB030907, AF238486, D10988; 2c: D50409; 2i: DQ155561; 2k.AB031663; 3a.X76918, AF046866, D17763; 3b.D49374; 3k.D63821; 4a.NC_009825, DQ418788; 4d.DQ418786, DQ516083; 5a.AF064490, NC_009826; 6a: DQ480513, Y12083; 6b: D84262; 6c: EF424629;

7a: EF108306. The specific reference strains used separately in the 2k/1b analyses were 2k/1b.AY587845 and 2k/1b.FJ821465; in the genotype 1 analyses are 1g.AM910652; and in the genotype 4 analyses were 4b.FJ025854, 4b.FJ025855, 4b.FJ025856, 4b.FJ462440, 4c. FJ462436, 4f.EF589160, 4f.EF589161, 4f.EU392169, 4f.EU392170, 4g.FJ462432, 4k.EU392171, 4k.EU392173, 4k.FJ462438, 4l.FJ839870, 4m.FJ462433, 4n.FJ462441, 4o.FJ462440, 4p.FJ462431, 4q.FJ462434, 4r.FJ462439, and 4t.FJ839869.

3.3. Results and discussion

3.3.1. Experimental fidelity

The experimental designs for the amplification and sequencing of the near-full genomes of 2k/1b strains, an unclassified genotype 1 isolate, and unclassified genotype 4 strains were successful. For the recombinant strains, due to genetic variations between the two samples, not all PCR and sequencing primers were appropriate for both, and so strain-specific primers were designed where necessary, resulting in two slightly different designs (Figure 8 for CYHCV037 and Figure 9 for CYHCV093) probably due to the sequence diversity between the strains, leading to varying efficiency or primer annealing. This highlights the problem that HCV genetic heterogeneity causes to the designs of molecular assays. For the genotype 4 samples, only two of the four candidates were successful in this application, both with the same protocol (Figure 11). The two that were PCR-negative were the samples that were sequenced in only one of the two genomic regions in the molecular epidemiology study. For all samples that were PCR positive in this design, overlapping PCR products could be obtained across the full genome, and the resulting sequence readings were clear, implying no mixed infections in the samples in question.

3.3.2. 2k/1b recombinant strains

During the study of the molecular epidemiology of HCV infection in the general population of Cyprus, two samples were found to be subtype 2k in the Core-E1 region, and 1b in the NS5B region (indicated with asterisks in Figure 5). The samples were taken from two Georgian males being seen at different public hospitals. A laboratory code was given to each sample; CYHCV037 and CYHCV093. Patient CYHCV037, 32 years old when sampled in 2005, had first been diagnosed HCV-seropositive in 2003, and stated sexual transmission as

the possible route of infection. Patient CYHCV093, 39 years old when sampled in 2007, had been diagnosed HCV-seropositive in 2000, and stated intravenous drug use as the transmission route. Neither individual had taken antiviral therapy at the time of sampling.

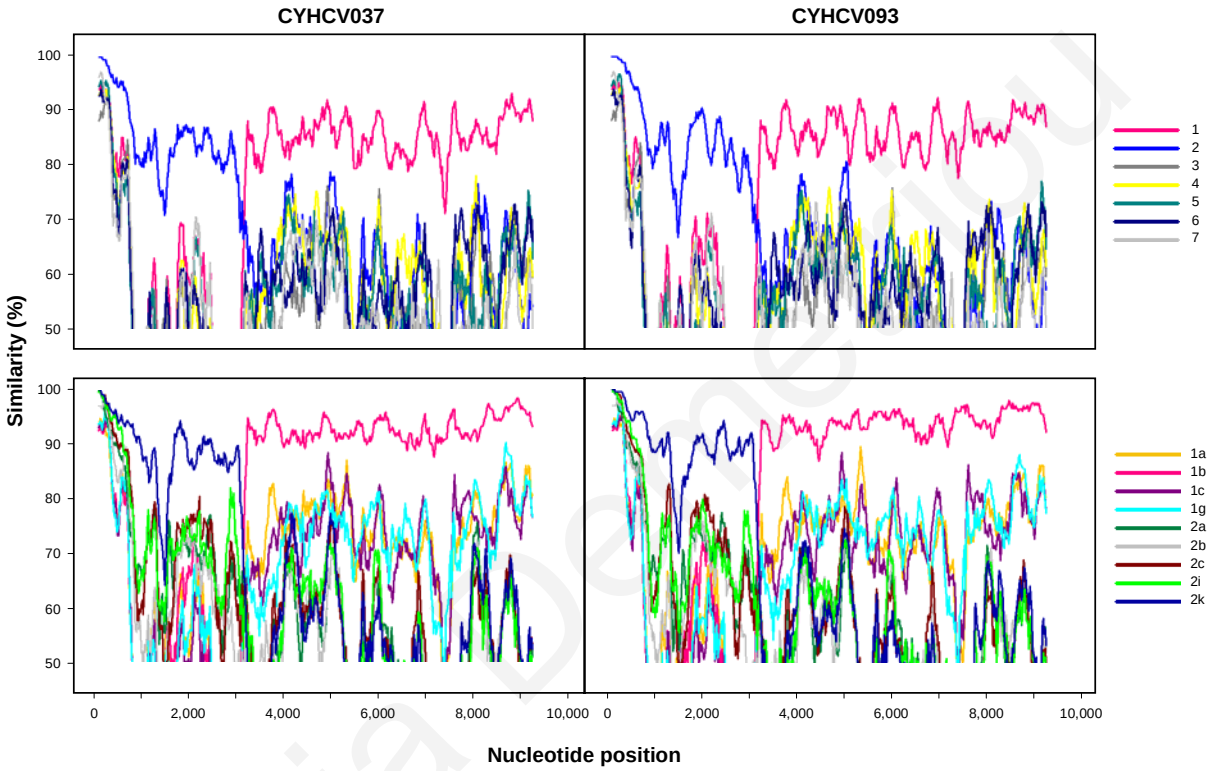

Figure 12. Simplot analyses of the two 2k/1b recombinant candidate strains, showing percentage sequence similarity over nucleotide position. The plots on the right are for strain CYHCV093, and on the left for strain CYHCV037. The plots on the top are the results of a genotype-level analysis, with genotypes corresponding to line colours annotated on the right of the plots. The bottom plots are the results of a subtype-level analysis, once the genetically closest parental genotypes had been determined using the top plots. The subtypes corresponding to the coloured lines are annotated on the right of the plots.

The near-full length sequences (nt. 87-9287, positions according to strain H77) were successfully derived from the two samples from 5-6 overlapping PCR fragments (Figure 8 and Figure 9). To characterise the sequences, compare them to classified reference strains and investigate the event of recombination along the length of the sequence, SimPlot and Bootscan analyses were carried out. The Simplot graph was constructed at first at the genotype level using consensus sequences derived from the references, and then, following determination of

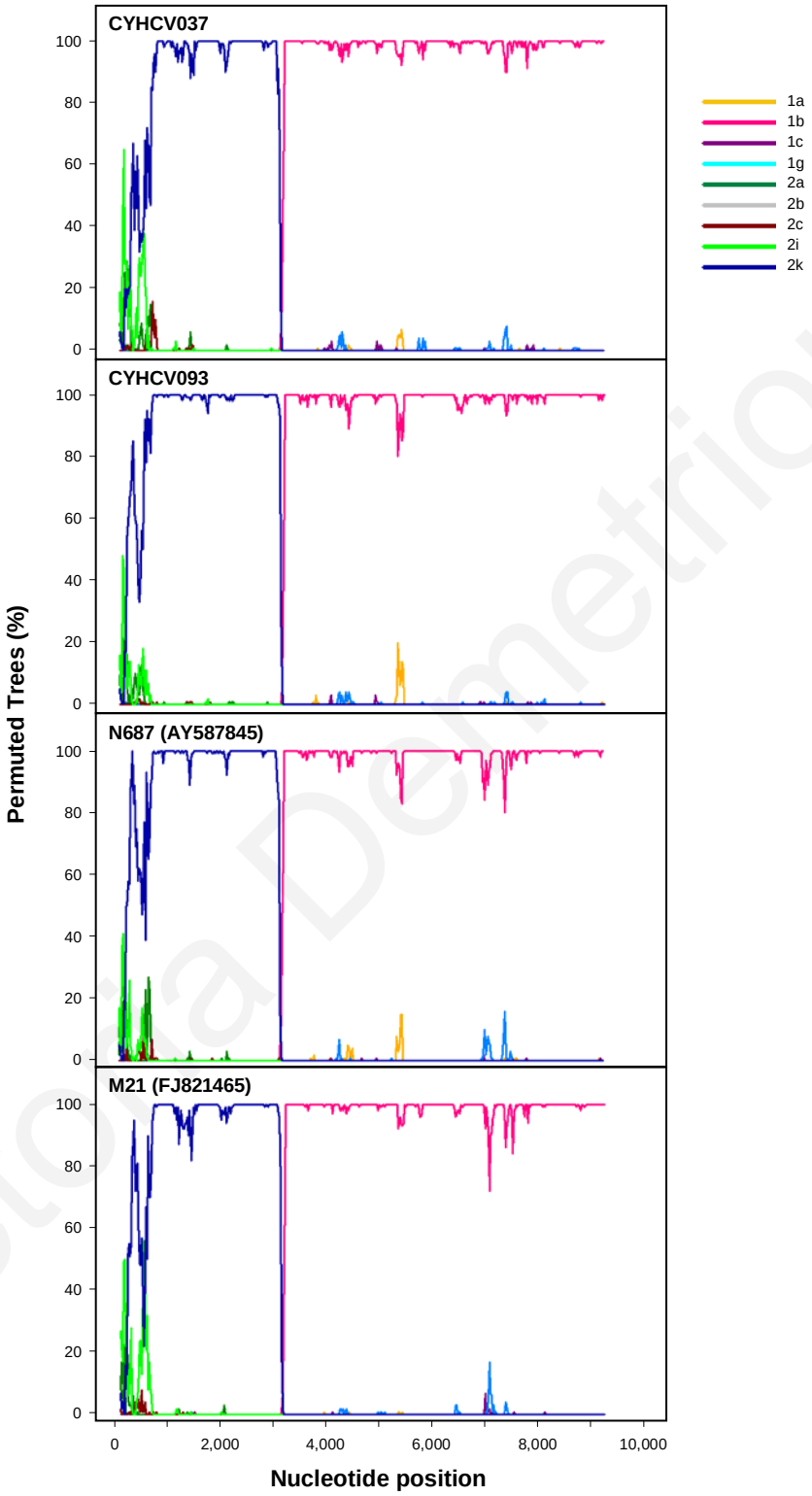

Figure 13. Bootscan plot of percentage permuted trees over nucleotide position across the full length of the 2k/1b recombinant strains. The top two plots are for the strains described in this study and the bottom two for the only near-full 2k/1b genomes submitted to GenBank to date. Strain identifiers are indicated at the top left corner of each plot. In brackets are the GenBank accession numbers of the publicly available sequences. The annotation on the right shows the HCV subtype of the consensus sequences corresponding to each line colour in the plots.

the parental genotypes, at the subtype level of the parental genotypes (Figure 12). Following the SimPlot analyses, Bootscan plots were constructed for each sequence using subtypes of parental genotypes 1 and 2 as references (Figure 13). For comparison, this was also done for the two already available full genomes of 2k/1b strains. From the results, it is clear that the strains identified in this study were more similar to genotype 2, specifically 2k, on the 5′ side of their genomes, and more similar to genotype 1, specifically 1b, on the 3′ side of their genomes. It is therefore confirmed that they are 2k/1b recombinant strains, and it is clearly seen that the point of recombination is in the NS2 region, after position 3,000. The Bootscan plots plainly demonstrate that the recombination point in the two strains found in Cyprus is in the same genomic location as the original 2k/1b strain, isolate N687 (Acc. No. AY587845), found in St. Petersburg and later characterised across the full genome (Kalinina *et al.*, 2004; Kalinina *et al.*, 2002), and a second 2k/1b recombinant, strain M21 (Acc. No. FJ821465), discovered recently in France (Morel *et al.*, 2010). The nucleotide and deduced amino acid sequences of the two CYHCV strains were according to the original Russian N687 strain, with no insertions or deletions noted. The aforementioned strains are the only two 2k/1b full

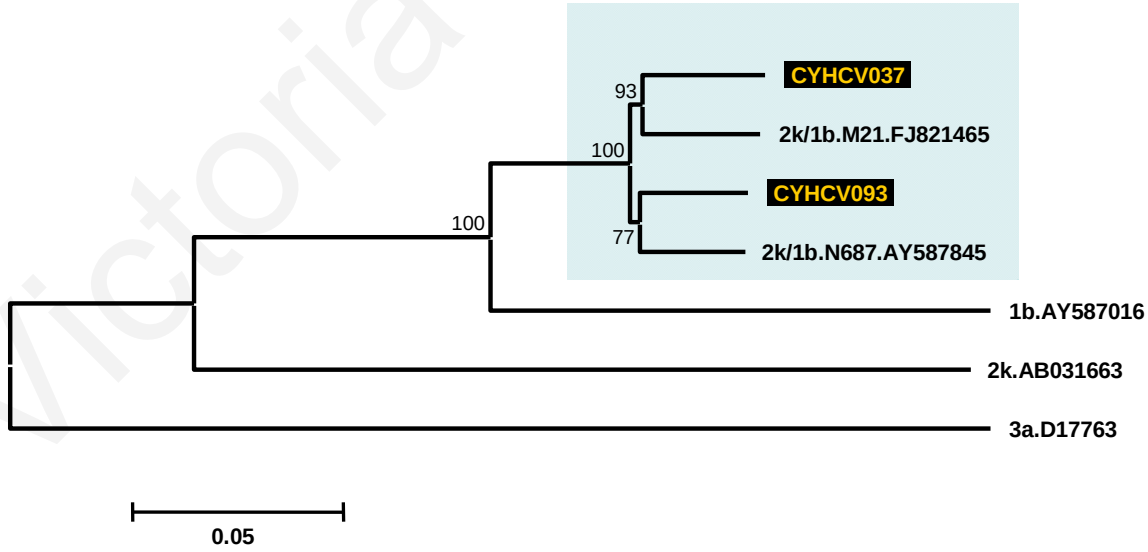

Figure 14. Neighbour-Joining tree showing the phylogenetic relationships of the two near-full length 2k/1b isolates described in this study, highlighted in black, the two 2k/1b strains isolated in Russia and France, and reference strains for parental genotypes 1b and 2k. Numbers at the branch nodes are percentage bootstrap support for 1,000 replicates. A genotype 3a strain is used as an outgroup. The scale on the lower left indicates the genetic divergence between any two sequences as obtained from the total branch length.

sequences currently available in GenBank and the results from this study demonstrate that the strains sequenced here are recombinants of the same type, increasing the current near-full genome sequence data for this strain by 100%.

Subsequently, phylogenetic analysis was performed on the CYHCV037 and CYHCV093 sequences in comparison to the available 2k/1b full genome sequences, N687 and M21, and reference strains representing the parental genotypes 1b and 2k. The resulting tree is seen in Figure 14. The 2k/1b strains are seen to group together with a bootstrap value of 100, but within this group, the four strains are separated into two subgroups, CYHCV093 with N687 from Russia and CYHCV037 with M21 from France, with high bootstrap values, 77 and 93, respectively. The divergence between the sequences was estimated from the pairwise distances, and the average number of base substitutions per site among the four recombinants was 0.061 (0.054-0.064). The genetically closest sequences are CYHCV093 with N687, and the most divergent are CYHCV037 with N687, but the difference between these divergences is negligible, as all four strains are genetically very close. Phylogenetic trees were also constructed for the regions on either side of the recombination point of these strains (Figure 15). The additional sequences derived from the HCVBLAST searches are AY070214 and AY070215, both 2k/1b strains, used in the tree of the 5′ region of the genome, and EU155337 and EU155333, both 1b, used in the tree of the 3′ segment. The partial regions of the two strains described here group within the 2k and 1b subtypes, respectively, with a bootstrap value of 100. The 2k/1b recombinants cluster together in both trees, within the subtype group, with bootstrap support >98, signifying an evolutionary separation from the pure subtypes. The 1b tree (Figure 15B) presents the same topology and similar statistical significance of grouping as the whole genome tree (Figure 14), whereas the topology of the 2k/1b cluster within the 2k subtype in Figure 15A is not the same, with any subgroups within it displaying very low bootstrap support (< 55). This difference in clustering behaviour between the regions on either side of the recombination point is interesting, and could possibly signify that in the original recombination event, the 2k strain was in the presence of a genetically diverse population of 1b, perhaps due to multiple points of infection (due to unsafe injecting drug use practises), or that the host was infected with a 1b subtype for a long enough time to establish a genetically diverse 1b quasispecies population, before acquiring the 2k infection. In these two possible scenarios, recombination events could have occurred between the 2k genomes and

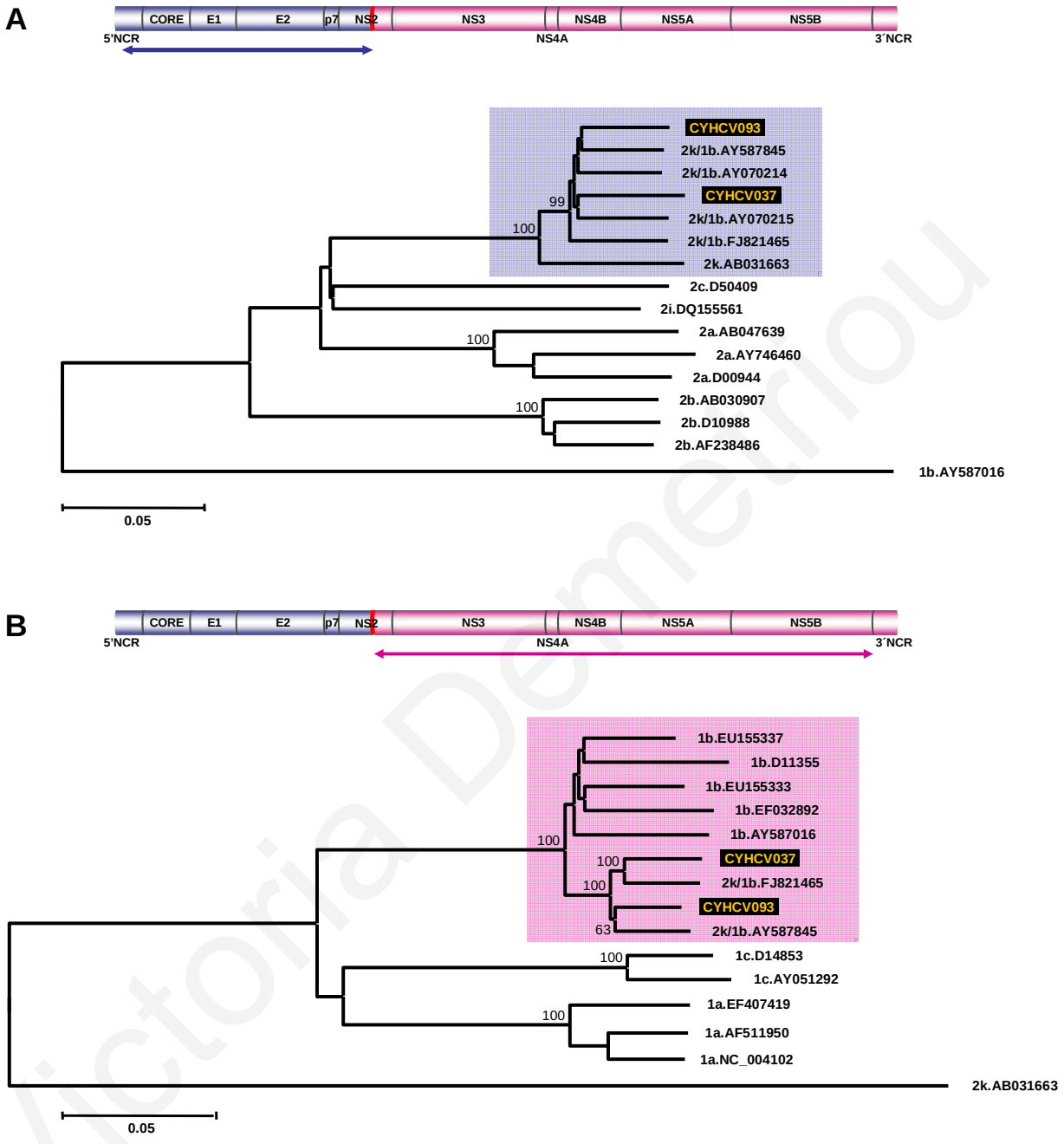

Figure 15. Neighbour-Joining trees for (A) the region 5' of the recombination point and (B) the region 3' of the recombination point, as seen by the indications on the genome maps above each tree, for the 2k/1b recombinant strains identified in this study (highlighted in black), other 2k/1b sequences obtained from GenBank, and reference stains of the corresponding parental genotypes. The trees are rooted with 1b and 2k strains, respectively. Grouping can be seen within the corresponding parental subtypes and is highlighted with coloured boxes. Numbers at branch nodes indicate the percentage bootstrap support for 1,000 replicates. The scale at the bottom left of each tree indicates the divergence between any two sequences obtained by summing the branch lengths between them.

genetically diverse 1b genomes within the host's environment, and the spread from two of these recombination events is represented in the strains that have been sequenced to date.

With the addition of more near-full genomes within this recombinant type, the evolutionary picture of this strain will become more apparent. With just one full genome available until this year (Kalinina *et al.*, 2004), and the recent submission of one more (Morel *et al.*, 2010), it has been impossible to draw any conclusions regarding its molecular evolution, but with the addition of two more near-full genomes from this study, hypotheses about this can begin to take form. From the topology of the tree and the bootstrap supports on the branch nodes, the recombinant demonstrates a biphyletic separation of its isolates (Figure 14). Since its discovery in 2002 (Kalinina *et al.*, 2002), only recently could evolutionary deductions begin to be proposed. A group in Uzbekistan carried out a molecular epidemiology investigation using partial sequences (Kurbanov *et al.*, 2008b), and they suggest that all the recombinant strains available to them share a common evolutionary ancestor, suggesting that intergenotypic recombination in HCV is sporadic, and not associated with co-circulation of different genotypes in a population. The authors observed that phylogenetically, the strains were tightly grouped together, forming distinct subclusters within the respective parental subtype groups on the Core-E1 and NS5B trees, suggesting that "the current 2k/1b viral population likely evolved from a single source introduction and subsequently spread rather than from a convergent evolution of distant strains". Kalinina *et al.* support this hypothesis, discussing that the probability of multiple recombination events involving the 2k subtype, considering its relatively low frequency and estimated rate of mixed infection, is very low (Kalinina *et al.*, 2002). However, from the data presented here, it can be supposed that the host who was the single source that introduced this strain to others was likely superinfected with different 1b strains, giving rise to a pool of recombinants. More near-full genome sequences of this recombinant are required for a more accurate evolutionary theory.

The natural 2k/1b recombinant was first identified in 2002 in St. Petersburg, Russia (Kalinina *et al.*, 2002) from five intravenous drug users, one of which was later sequenced along the full open reading frame (isolate N687) (Kalinina *et al.*, 2004). The recombination point was mapped to be between nucleotide positions 3164 and 3165 (corresponding to strain H77) within the NS2 region of the genome. It is thought to have emerged by homologous recombination during minus-strand synthesis via template switching (Kalinina *et al.*, 2004) or by "strand switching" during replication (Moreau *et al.*, 2006). In 2006, two isolates of this

recombinant were reported in Ireland, in individuals originating from Russia and Georgia (HC9A98987, Acc. No. AB327017 and AB327057; HC9A99966, Acc. No. AB327018 and AB327058) (Moreau *et al.*, 2006). Another such isolate was also discovered soon after in Estonia (a former member of the Soviet Union) in a 21 year old male patient sampled in 1998, who had had multiple blood transfusions (2032-98, Acc. No. EF194953) (Tallo *et al.*, 2007). In 2008 a study was published where one more strain (UZ-IDU19, Acc. No. AB327120-2) was found in a 32 year old male intravenous drug user, native of Uzbekistan (Kurbanov *et al.*, 2008a). Six new RF1 cases were later identified, five in South Siberia, Russia (strains ALT30, female, 34; KNG318, male, 30; KNG327, male, 25; PSA108, male, 37; and PSA62, male, 50) and one in Uzbekistan (strain AZ15, male, 22) (Kurbanov *et al.*, 2008b), whose modes of transmission were blood transfusions, injecting drug use, or unknown. Additionally, three more partial sequences of 2k/1b isolates were directly submitted to GenBank in 2004-2005 from samples taken in South Siberia (strains HIA1002, Acc. no. DQ001221 and DQ001269; KNG318, Acc. no. AY764169-72; and KNG327, Acc. no. AY764173-76). The most recent case of this recombinant was found in France, (M21, Acc. no. FJ821465), taken from a 30 year old Georgian male who was an intravenous drug user co-infected with HCV genotype 3a (Morel *et al.*, 2010). This and the original N687 strain are the only ones that have been sequenced along the complete genome so far. The demographic characteristics of these cases and the ones described in this study demonstrate a trend pointing towards individuals of nationality associated with the former Soviet Union, a relatively young age, and injecting drug use as the main risk behaviour. Hence, of the different types of recombinants found, the 2k/1b type appears to be circulating in small populations more than any other, as several infections have now been reported. The cases found indicate that the variant is circulating among intravenous drug users within countries of the former Soviet Union, and through migration, is spreading to other European regions.

Therefore, epidemiologically, HCV recombinant 2k/1b was initially identified in Eastern Europe and Central Asia and appears to be spreading in Europe and the Mediterranean region, through population movements and intravenous drug use. These findings demonstrate that this recombinant is the most widespread and broadly distributed for HCV, highlighting the need to investigate recombination in epidemiological studies, and to further investigate its response to antiviral therapy. As genotype 1 exhibits the lowest sensitivity to therapy, recombinants with this genotype as part of their genome (in particular the part of the genome

containing the ISDR within NS5A) may prove problematic in relation to response to interferon-based treatment. One case study of a patient infected with 2k/1b demonstrated viraemic re-emergence after therapy (Morel *et al.*, 2010) and recent data on the susceptibility of a 2k/1b variant to interferon in a chimeric mouse model demonstrated a good therapeutic response (Kurbanov *et al.*, 2008b). With the continual emergence of new 2k/1b recombinants, more in-depth studies need to be carried out to determine their sensitivity to the currently available therapeutic regimen.

The serendipitous identification of two recombinant forms of HCV, related to the known 2k/1b recombinant form, in Cyprus was the result of a molecular investigation across the genome. These findings highlight the opinion that the actual prevalence of this HCV type has been underestimated. The study shows that the 2k/1b recombinant is more widely distributed than was originally estimated, and that its actual molecular evolution is possibly more complicated than the hypothesis of one recombination point giving rise to one single evolutionary ancestor.

3.3.3. Unclassified genotype 1 strain

One strain in the general population group was found to belong in the genotype 1 group, but did not demonstrate phylogenetic association with any known subtypes of this group (Figure 6A & B). The sample was taken from a 48 year old Cypriot male (lab code CYHCV025), who had been diagnosed with HCV for three years at the time of sampling. The patient was on a course of pegylated interferon and ribavirin when blood was taken, after a relapse following a course of therapy two years before. The route of HCV transmission in this case was not known by the participant, therefore no particular risk behaviour or geographic origin can be associated with this new strain.

The near-full genome sequence (nt. 107-9293, positions according to strain H77) of the unclassified strain, CYHCV025, was successfully determined through RT nested PCR and sequencing (Figure 10) from four overlapping PCR fragments. The sequence was characterised by means of Simplot, Bootscan and phylogenetic analyses. The Simplot and Bootscan results are seen in Figure 16. The data, using reference sequences available from GenBank, show that the sequence of CYHCV025 does not appear to be a recombinant of any known genotypes. From an analysis made at the genotype level (Figure 16A), the strain found here appears to be genetically closest to genotype 1 across its whole nucleotide sequence. The

sequence similarity to genotype 1 after the 5´NCR, which is the most conserved region of the HCV genome, ranges from ~55-85%. Following this observation, a second analysis was carried out on a subtype level, where strains of all available genotype 1 subtypes were compared to the CYHCV025 sequence (Figure 16B), showing that the strain in question is equally genetically distant to them. No specific subtype prevailed or demonstrated a percentage sequence identity of >85% across the length of the sequence. A Bootscan analysis showed that certain small stretches of the sequence were genetically closer to subtypes 1b and 1g than any other, but the lengths of the nucleotide segments that showed percentage permuted trees >80% were very small (< 300 nt).

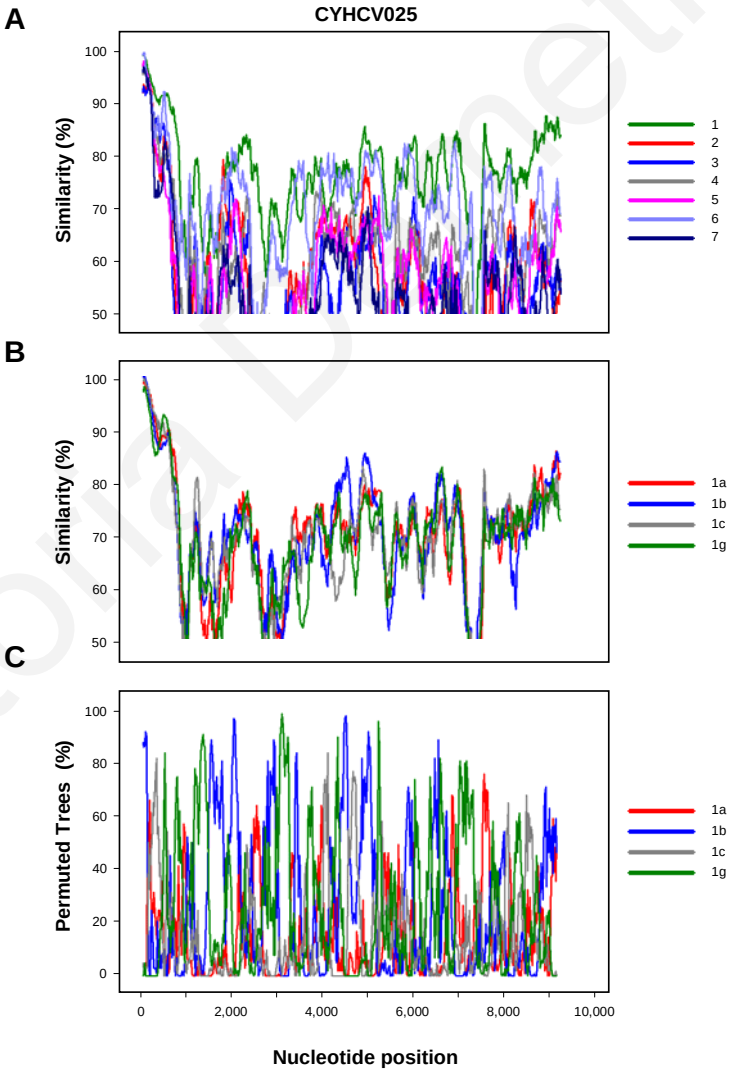

Figure 16. Analysis of strain CYHCV025 across its full length sequence. (A) Simplot results of comparison with genotype consensus strains. (B) Simplot results of comparison with genotype 1 subtype consensus strains. (C) Bootscan plot of analysis against genotype 1 subtype consensus strains.

A phylogenetic tree was also constructed using a near-full genome alignment of the strain described here and reference sequences representing all HCV genotypes and all available genotype 1 subtypes (Figure 17), showing the relationships and genetic distances between them. The tree confirms that the CYHCV025 isolate is more closely related to genotype 1 than any other genotype, and groups closely to this genotype cluster with a bootstrap support of 100, distinct from any subtype groups. Also, it appears to be more genetically distant from the genotype 1 subtypes than they are from each other. The average pairwise distances per site calculated between CYHCV025 and the various genotypes are: 0.309 from genotype 1, 0.518 from genotype 2, 0.504 from genotype 3, 0.418 from genotype 4, 0.461 from genotype 5, 0.464 from genotype 6, and 0.524 from genotype 7. From known genotype 1 subtypes, the strain described here exhibits the following numbers of base substitutions per site: 0.304 from 1a, 0.308 from 1b, 0.317 from 1c, and 0.311 from 1g, showing no significant similarity to any particular subtype over another. In fact, the percentage genetic variability from genotype 1 (~31%) implies that this strain is on the verge of belonging to a different, as yet undiscovered, genotype altogether (Simmonds *et al.*, 2005), and is simply closer to genotype 1 than the other genotypes, from which it differs by 41-52% across the near-complete genome sequence. It should therefore be referred to as a genotype-1-like strain. This unclassified strain could represent a new genotype 1 subtype, or even a new HCV genotype, but more epidemiologically distinct strains that cluster with it are needed to assign a new subtype or genotype; such sequences were neither found in this study nor in the HCV sequence database. If more such strains are identified, this isolate could then be classified and would perhaps be the first isolate of the next genotype 1 subtype.

Overall, the near-full genome sequence of a novel HCV strain is presented here, that does not classify within any known HCV subtypes. From the results of comparison with the current available sequence data, there is no reason to believe that this strain is a recombinant, but it is more likely a subtype of genotype 1 which has not been discovered until now. It could represent a putative novel subtype, but confirmed designation cannot be made, as no more examples of this strain have been found either in this study or as submissions to the HCV database. This is the first report of this subtype in either partial or full genome sequence data. Sequence comparison analyses reveal that the strain is closer to genotype 1 than any other genotype but its genetic diversity from available classified sequences suggests that it is on the

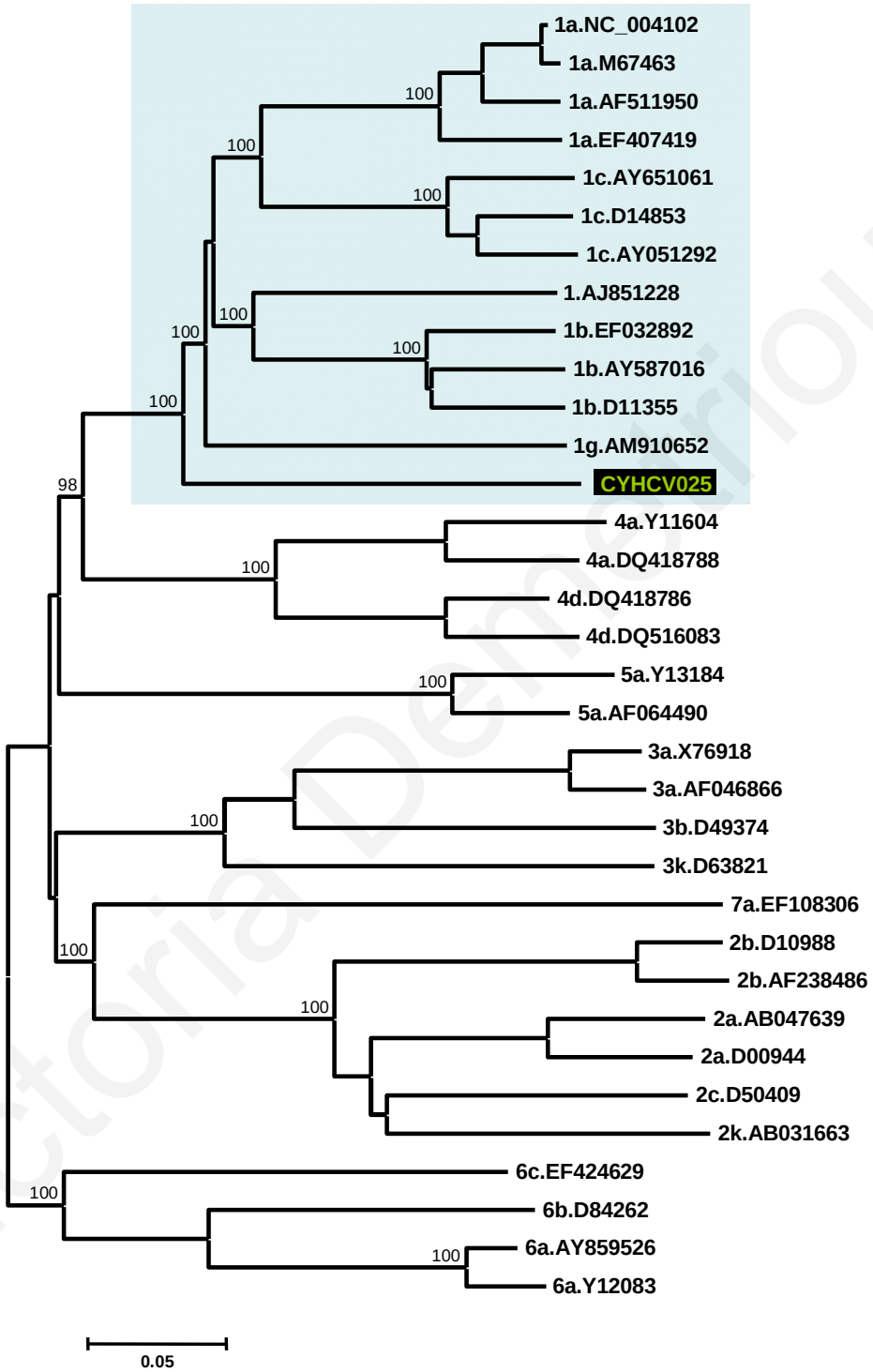

Figure 17. Neighbour-joining tree of near-full genome sequences showing the classification of strain CYHCV025 in genotype 1 (within the blue box), but not in any known subtype. Strain CYCHV025 is highlighted in black and reference sequences are written in black, indicating their subtype and GenBank accession number. Numbers at branch nodes signify the percentage bootstrap support for 1,000 replicates. The genetic distance relative to branch length between any two sequences is denoted by the scale on the bottom left.

verge of being a new HCV genotype. As is appears to be genetically close to genotype 1, it is interesting, as most new HCV subtypes identified recently belong to genotypes 4 and 6. This highlights the continuous evolution of this virus and signifies its ability to change and manipulate itself genetically to adapt to its host and environment. The existing lack of full or near-full genome sequence data publicly available for non-a, non-b subtypes of genotype 1, and the low intra-subtype variation within this genotype, present a significant limitation in classifying the variant described here.

3.3.4. Unclassified genotype 4 strains

Isolates of a third type of strain identified in the general population cohort were considered worthy of near-full genome sequencing and analysis. Two samples were sequenced in both genomic regions investigated, one in Core-E1 only, and one in NS5B only. These sequences grouped within genotype 4 and formed clusters with each other, distinct from any known genotype 4 subtype with available sequences in the genomic regions investigated (Figure 6C & D), and were therefore considered to possibly belong to a novel genotype 4 subtype. The samples which were PCR-positive in both regions were isolated from an 84 year old Cypriot male (lab code CYHCV048), infected via blood transfusion, and diagnosed in 2000, and the other from a 71 year old Cypriot male (lab code CYHCV073), who was diagnosed in 1996, had been on multiple courses of therapy, but did not know how he had been infected. The strain positive only in the Core-E1 region was isolate CYHCV081, from a 58 year old female Greek patient born in Sudan, diagnosed in 1993, and whose stated transmission route was an unsafe surgical procedure. Finally, the strain positive only in the NS5B region was isolate CYHCV126, from a 33 year old Bulgarian woman, who had been on a course of therapy. No further epidemiological details are known for this patient.

The near-full genome sequences (nt. 87-9271, positions according to strain H77) were successfully derived from two of these samples, CYHCV048 and CYHCV073, from four overlapping PCR fragments (Figure 11). They were subsequently included in Simplot, Bootscan and phylogenetic analyses for sequence characterisation. As with the strains described in the sections above, these sequences were analysed at the genotype level before proceeding to more specific comparisons at the subtype level. The Simplot graphs of both isolates are seen in Figure 18, showing initially that the percentage sequence similarity along the whole derived sequence is significantly higher with genotype 4 than any other genotype,

by comparison to consensus sequences of reference genotype strains available from GenBank. At the subtype level, the strain was compared to near-full genome sequences of all available genotype 4 subtypes. For many, only one full genome sequence is available, but for the rest, a consensus sequence was made automatically in the Simplot software using the available assigned sequences from the input alignment. In the resulting plots (Figure 18B), all subtypes demonstrate a similar percentage sequence similarity, with 4q being on the highest end of the spectrum. The Bootscan plots revealed a significantly higher percentage of permuted trees supporting similarity of the strains described in this study with subtype 4q, than with any other subtype. These percentages however, are mainly lower than 70% for the largest part of the sequences, and therefore do not support that the strains belong in this subtype. The results of these analyses do not reveal any recombination evens within the sequences. The Simplot and Bootscan graphs of the two Cypriot strains are very similar, implying that they are genetically close to each other. Therefore, these seemingly unlinked patients probably share a common risk factor, such as percutaneous exposure to infected blood while travelling in Central Africa, and or infection from each other when in Cyprus, but such epidemiological data is missing.

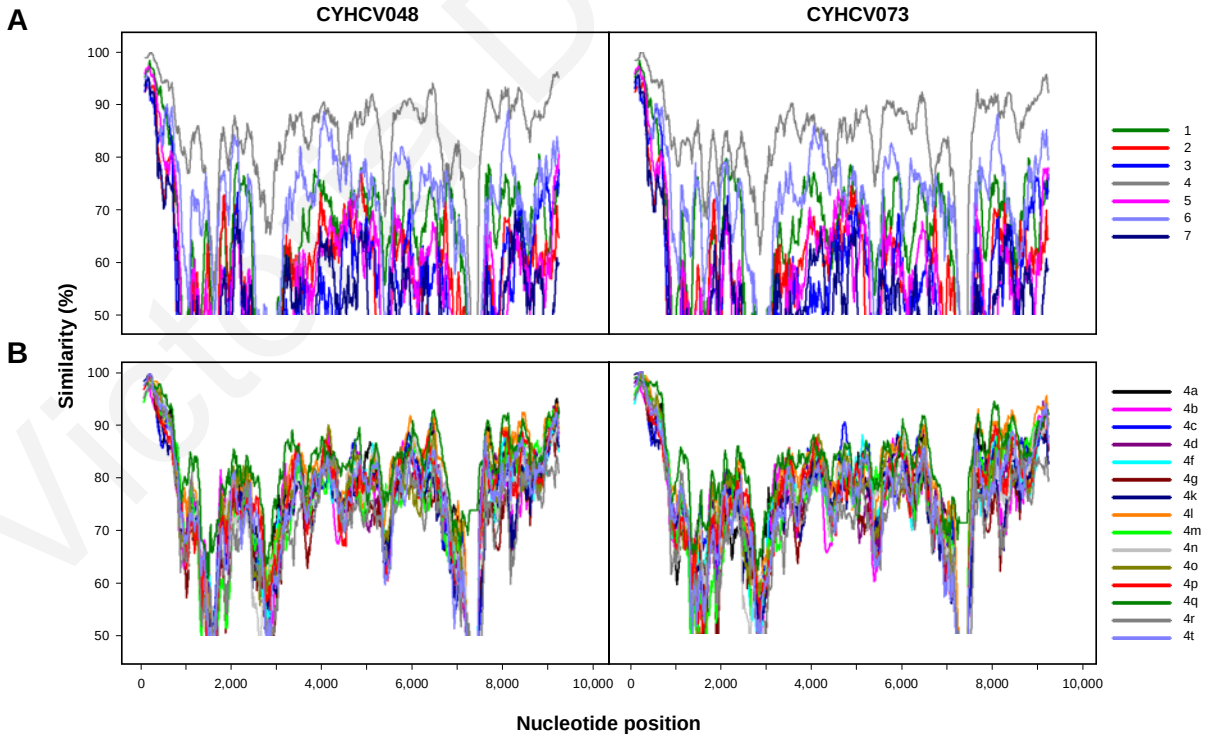

Figure 18. Simplot analysis across the full sequence of unclassified genotype 4 strains CYHCV048 and CYHCV073 showing percentage sequence similarity over sequence position. (A) Analysis on a genotype level by comparison to genotype consensus sequences. (B) Analysis on a subtype level by comparison to genotype 4 subtype consensus sequences.

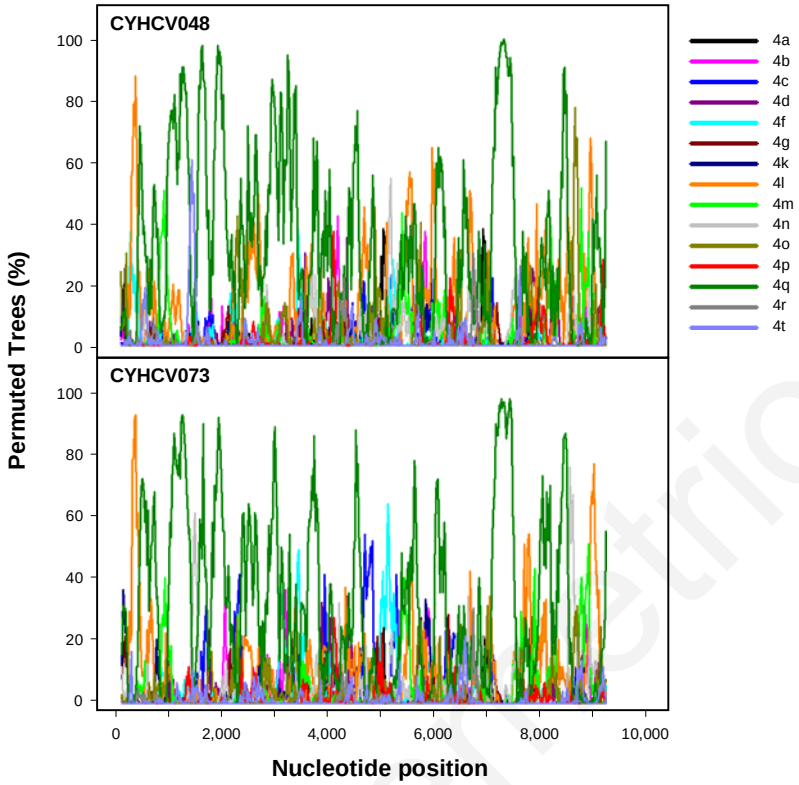

Figure 19. Bootscan plots of percentage permuted trees across the full sequence length for unclassified genotype 4 strain CYHCV048 (above) and CYHCV073 (below). The annotation on the left shows the subtypes corresponding to the coloured lines in the plot.

The near-full genome sequences of the strains were then included in a phylogenetic analysis after alignment with near-full genomes of reference sequences representing all HCV genotypes and all available genotype 4 subtypes (Figure 20). The resulting tree shows that the Cypriot strains definitely belong in genotype 4, they cluster together with a bootstrap support of 100, and are genetically closest to subtypes 4q and 4l. In fact, the 4l and 4q strains group with the CYHCV048 and CYHCV073 isolates with a bootstrap support of 100, and within that group, the Cypriot strains form a subgroup with the 4q sequence with a high bootstrap support (94). The estimated average pairwise distance between the two Cypriot strains is 0.059 base substitutions per site. The pairwise distances calculated between the strains found here and all HCV genotypes and genotype 4 subtypes are seen in Table 7. The genetic variation from sequences of genotype 4 is small enough to classify them within this genotype. The evolutionary distance from the various genotype 4 subtypes shows that they are close to 4q

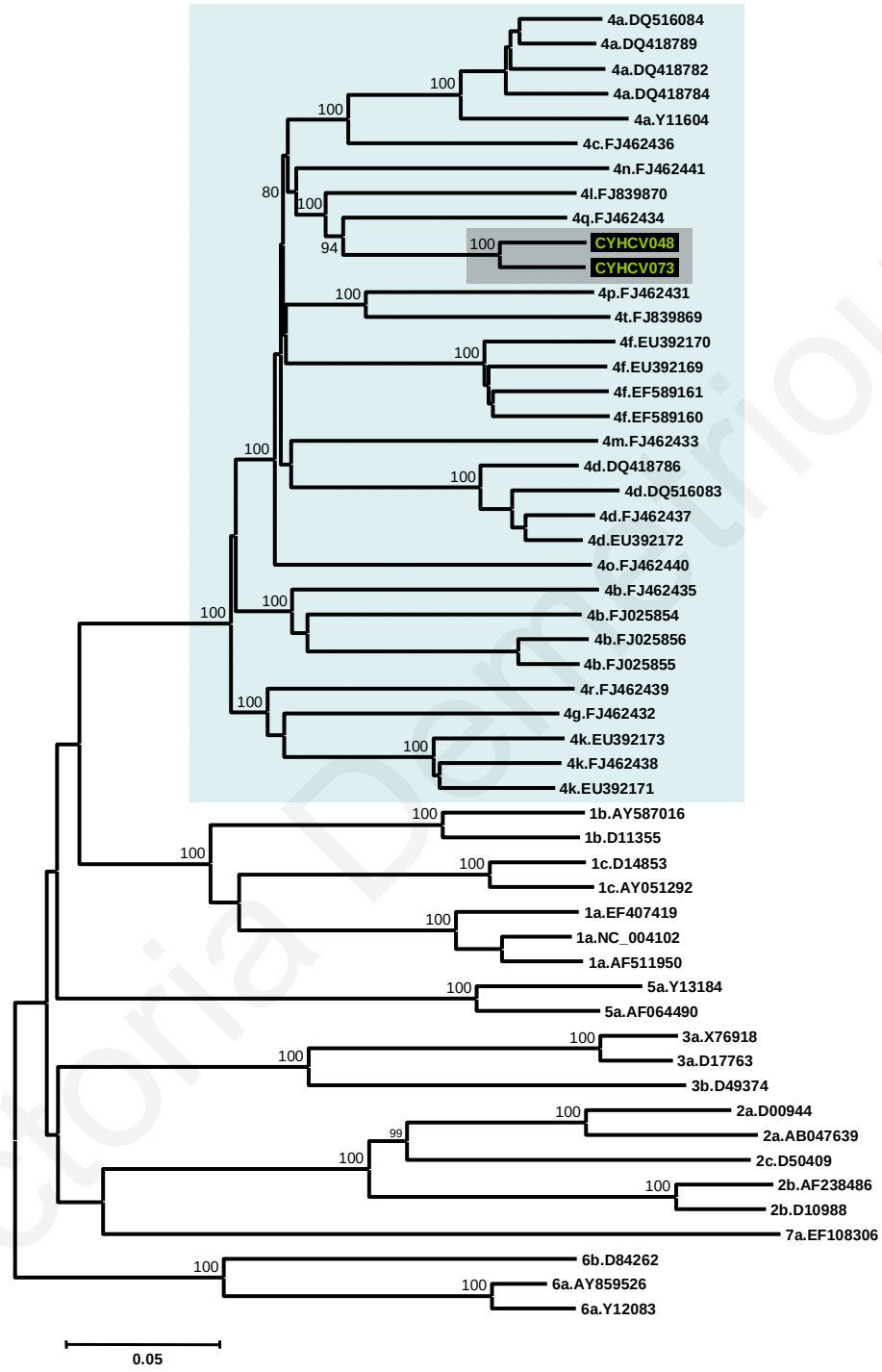

Figure 20. Neighbour-joining tree of near-full genome sequences including the unclassified genotype 4 strains described in this study (highlighted in a grey box), which can be seen to classify in the genotype 4 group (within the blue box), but not within any known subtype. Reference strains used to construct the tree are written in black type, indicating subtype and GenBank accession number. Numbers at the branch nodes denote bootstrap values; only those above 70 are shown. The branch length per unit genetic distance is indicated on the scale at the bottom left.

and 4l (17% and 19%, respectively), and on the verge of belonging to these subtypes. In fact, as 4q and 4l demonstrate 18% variation across the near-full genome, and given the deeper bootstrap support on the phylogenetic tree, all of these strains could be grouped in one subtype according to the established criteria for classification (Simmonds *et al.*, 2005), and for the purposes of simplifying HCV nomenclature. However, in the most recent publication describing the determination of the full genome reference sequences for these subtypes, the authors state that: "Although these observed similarities were higher than those suggested by the updated HCV nomenclature (75.0-80.0%), for subtypes of a single genotype (Simmonds *et al.*, 2005), these isolates represented 13 distinct genotype 4 subtypes." (Li *et al.*, 2009a). This genetic similarity between subtypes may indicate a continual genetic variation of HCV with many variants still missing, and some closely related subtypes may qualify as a single subtype if such missing variants are found. Due to the ever increasing number of HCV subtypes, it is important to be cautious before assigning new subtypes to newly discovered isolates. As more HCV variants are identified, new data will be obtained and can be used to develop additional nomenclature and classification regulations.

Table 7. The number of base substitutions per site between near-full genome sequences.

	Genotypes						
Isolate	**Gen 1**	**Gen 2**	**Gen 3**	**Gen 4**	**Gen 5**	**Gen 6**	**Gen 7**
CYHCV048	0,412	0,539	0,498	0,240	0,454	0,468	0,548
CYHCV073	0,409	0,545	0,491	0,240	0,457	0,466	0,547

	Genotype 4 subtypes							
	4a	**4b**	**4c**	**4d**	**4f**	**4g**	**4k**	**4l**
CYHCV048	0,238	0,264	0,222	0,235	0,243	0,258	0,258	0,193
CYHCV073	0,236	0,270	0,222	0,233	0,243	0,252	0,259	0,194

	4m	**4n**	**4o**	**4p**	**4q**	**4r**	**4t**
CYHCV048	0,235	0,229	0,229	0,232	0,171	0,279	0,234
CYHCV073	0,235	0,227	0,231	0,231	0,173	0,276	0,237

As discussed in section 2.3.3.1, the partial sequences of the Cypriot strains discussed here form a monophyletic cluster with certain other non-subtyped genotype 4 strains derived from the HCV database (Figure 6C & D). In the Core-E1 region, these are seven strains, four

from subjects of African origin from a Canadian submission (Acc. No. EF115885, EF115899, EF115905, and EF115910) (Murphy *et al.*, 2007b) and three from a direct UK submission (Acc. No. AY766949, AY767036, and AY767953) from patients of unknown epidemiological and demographic details, apart from that one of them may be of Middle Eastern origin (Dr. Phil Bennett, direct communication). A similar pattern is revealed in the NS5B tree, where the two sequences from this study cluster with the same four Canadian strains (Acc. No. EF116108, EF116122, EF116128, and EF116133), separately from any assigned strains. From the data available and associated with these strains, this cluster probably evolved from a common ancestor originating in Africa, but no specific risk behaviour can be linked to it yet. Also, it is interesting to note that is has been found in three very different geographical locations (Cyprus, Britain and Canada), with no evidence of how it has circulated. From the partial sequences, the topological closeness of this cluster to 4l and 4q is not supported by high bootstrap values.

HCV genotype 4 is prevalent in the Middle East and Africa, where it exhibits great genetic diversity as it has been endemic in these regions for a long time, and although it has a low prevalence in the Western world, it has recently been shown to be spreading to Europe and America due to immigration and intravenous drug use (Ansaldi *et al.*, 2005; Fernandez-Arcas *et al.*, 2006; Franco *et al.*, 2007; Kamal and Nasser, 2008; Katsoulidou *et al.*, 2006; Lyra *et al.*, 2004; Murphy *et al.*, 2007b; Nicot *et al.*, 2005; Payan *et al.*, 2005; Savvas *et al.*, 2005). According to the public databases, to date, 19 different subtypes (4a, 4b, 4c, 4d, 4e, 4f, 4g, 4h, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r, 4s, 4t and 4u) have been identified in different geographical regions (Hmaied *et al.*, 2007; Kamal and Nasser, 2008; Kuntzen *et al.*, 2008b; Li *et al.*, 2009a; Simmonds *et al.*, 2005; Timm *et al.*, 2007), demonstrating its widespread genetic heterogeneity, due to its endemic persistence in Central Africa for a long period of time. Phylogenetic analysis of the complete genome is important to ensure the correct classification proposed by Peter Simmonds and colleagues (Simmonds *et al.*, 2005).Until recently, the entire genome was available for only four of these subtypes (4a, 4d, 4f and 4k). Last year, complete genome reference sequences were published for subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t (Koletzki *et al.*, 2009; Li *et al.*, 2009a). The classification of these recognises the epidemiological associations these variants and their subtype names have been incorporated into the HCV nomenclature, but it has been noted that such labels are of little value in the

description of HCV variants in high diversity areas such as sub-Saharan Africa (de Bruijne *et al.*, 2009b; Kuiken and Simmonds, 2009; Simmonds *et al.*, 2005).

In summary, strains of an unclassified genotype 4 variant have been identified in this study and two of these have been sequenced along the near-full genome. Partial sequences of this new variant have been found elsewhere, but have not been classified within a subtype nor have their full genomes been sequenced to date. From the sequence analyses described above, these strains seem to be on the verge of belonging to a new genotype 4 subtype, or represent a monophyletic cluster within another subtype that includes the assigned 4q strain. Ideally, more genotype 4 strains genetically close to subtypes 4q and 4l need to be characterised in order to accurately classify these subtypes according to the established principles of HCV nomenclature and taxonomy (Simmonds *et al.*, 2005). Based on the phylogenetic analyses, the genetic variations from other subtypes, and the current picture of subtype classification in HCV genotype 4, it is proposed that this group of strains should be designated as 4v, as more than three epidemiologically distinct infections have been found (four in Canada, three in the UK, and four in Cyprus), and here, we present two near-full genome sequences. The sequences presented here are the first reference sequences for this variant, which can be used as references for HCV classification, but also for diagnostic and clinical purposes.

3.3.5. Nucleotide sequence accession numbers

The near-full length genome sequences of the uncommon HCV variants described here have been submitted to GenBank, and the accession numbers assigned to them are HQ537005 for 2k/1b strain CYHCV037, HQ537006 for 2k/1b strain CYHCV093, HQ537007 for strain CYHCV025, HQ537008 for genotype 4 strain CYHCV048, and HQ537009 for genotype 4 strain CYHCV073.

4. General Discussion and Concluding Remarks

Knowledge of the HCV genotype is not only important epidemiologically, but also essential from the clinical point of view (Pawlotsky, 2003c; Zein, 2000). HCV classification and nomenclature is now carried out under specific standardised criteria, and novel subtypes continue to be identified around the world (Simmonds *et al.*, 2005). In this thesis, investigations involving the sequence of the highly genetically variable hepatitis C virus are described from Cyprus, divided into two major sections.

In the first part of this study, the molecular epidemiology of the HCV infection in Cyprus is presented for the first time, through studies in the general population and high-risk cohorts, including intravenous drug users, inmates from the state correctional facilities and HIV/HCV coinfected individuals. To do this, samples were collected through collaborations with an extensive network of clinicians in private clinics and public hospitals. The samples were investigated using an RT-PCR and sequencing protocol designed to amplify two partial regions on either end of the HCV genome, Core-E1 and NS5B. It is worth noting that the high nucleotide sequence heterogeneity in HCV hampers the detection of some strains due to inefficient primer annealing. Identified strains therefore depend on primer design and experimental methods. However, the experiments here were designed in order to detect all HCV genotypes, and the protocol proved adequate for detecting strains from samples with detectable viral loads, for amplifying highly genetically different variants (genotypes 1-5), for subtype designation and phylogenetic analysis, and for detecting (but not confirming) recombination events.

From the general population an epidemiologically diverse set of samples was included with regards to country of origin and routes of transmission, representing the general picture of HCV infection on the island. Phylogenetic analyses of the retrieved sequences revealed a high genetic heterogeneity, broad genotype distribution, and a polyphyletic infection. The subtypes found in the Cypriot population (1a, 1b, unclassified-1, 2c, 2k/1b, 3a, 4a, unclassified-4 and 5a) are similar to the reported distribution in Greece (Katsoulidou *et al.*, 2006), but not similar to other countries of this geographical region (Abdel-Hamid *et al.*, 2007; Altindis *et al.*, 2006; Altuglu *et al.*, 2008; Ramia and Eid-Fares, 2006; Watson *et al.*, 1999), indicating the strong association of the island with Greece and the mixing of the two populations. Even though Cyprus is a small country, it is in a strategic geographical position in

between Europe, Africa and Asia, and these results highlight the influx of immigrants on the island (also due to the recent admission of the country into the European Union), demonstrating that this movement plays a significant role on the status of the HCV infection. Investigations on the subtype associated with the most high risk behaviour, injecting drug use, 3a, confirmed that this subtype showed no country-specific clustering, reflecting the very large scale at which injecting drug use occurs in Europe, and supporting a unique origin of this subtype.

The investigation of the HCV molecular epidemiology among intravenous drug users seeking rehabilitation in Cyprus found only genotypes 1b and 3a, and among inmates from the state penitentiary units, genotypes 3a, 1b and 1a were found. Subtypes 3a and 1a are strongly associated with this risk behaviour, but 1b has also been reported to circulate in this population (Mathei *et al.*, 2005; Oliveira *et al.*, 2009; Peng *et al.*, 2008; Tallo *et al.*, 2007; Zhou *et al.*, 2006). Even though an increasing incidence of genotype 4 has been reported in Europe, especially in Southern Europe (Ansaldi *et al.*, 2005; Fernandez-Arcas *et al.*, 2006; Katsoulidou *et al.*, 2006; Nicot *et al.*, 2005), no such cases were found here, even though they were present in the general population. Phylogenetic analysis of the sequences obtained from these groups confirmed the polyphyletic nature of the local epidemic, but certain small monophyletic clusters also appeared among them, suggesting a history of equipment sharing and unsafe injecting practises among small groups of users on the island. No strains from the general population were included in the monophyletic clusters with the high-risk groups, suggesting no spill-over between the two populations, and illustrating that the HCV epidemic in the high-risk groups runs in parallel to the general population infected with HCV. These small clusters are formed between strains from individuals who have common characteristics (nationality, age group, duration of injecting drug use) and injecting drug use is the common route of infection for all, highlighting the issue of rapid and uncontrolled transmission among these individuals.

The main risk factor associated with these high-risk populations was the duration of intravenous drug abuse, confirming results of other studies (Hagan *et al.*, 2007; Shepard *et al.*, 2005; Sweeting *et al.*, 2009). Even though intravenous drug use is generally considered to be the main risk factor for prisoners, the routes of transmission could be many, due to the range of high-risk behaviour in prisons, such as injecting drug use, unsanitary tattoo and piercing procedures and getting into fights with exposure to blood. The prevalence of HCV infection among the injecting drug user population was 50%, comparable to the rest of the world but

also significant enough to raise an alarm for the requirement of harm reduction strategies to be applied for the protection of intravenous drug users. The prevalence of HIV and HBV found in this group was zero. These viruses are known to be less prevalent than HCV in this type of cohort, as fewer sharing partners are necessary to sustain HCV transmission than for other blood-borne viruses, and indirect drug sharing and preparation practices have also been associated with HCV transmission (Alter, 2007; Crofts *et al.*, 1999; Miller *et al.*, 2009b; Tan *et al.*, 2008). This is considered to apply to this population since the percentage of injecting drug users requesting therapy which was sampled in this study was high (87%), and is considered a representative sampling group.

The results from the individuals coinfected with HIV and HCV showed the existence of HCV genotypes 1a, 1b and 3a, which are all genotypes previously found among coinfected subjects, but in very different geographical locations (Liu *et al.*, 2008; Ponamgi *et al.*, 2009; Pukhrambam *et al.*, 2007; Re *et al.*, 2008; Urbanus *et al.*, 2009; van Asten *et al.*, 2004; Xia *et al.*, 2008b). The three high-risk groups display a similar pattern of genotype distribution, due to the common risk behaviours practiced among them. HIV-HCV coinfection has been associated with injecting drug use and MSM behaviour, although even among coinfected MSM, injecting drug use is significantly associated with infection, as well as multiple other risk behaviours (Danta *et al.*, 2007; Ghosn *et al.*, 2008; Liu *et al.*, 2008; Mboto *et al.*, 2006; Pando *et al.*, 2006; Ponamgi *et al.*, 2009; Pukhrambam *et al.*, 2007; Re *et al.*, 2008; Rhodes *et al.*, 2005; Tan *et al.*, 2008; Urbanus *et al.*, 2009; van Asten *et al.*, 2004; Xia *et al.*, 2008b). Few molecular epidemiology studies have been carried out in HIV/HCV coinfected populations in the world, and so little data exists to see specific patterns of genotype distribution and sequence variation among the HCV strains circulating in individuals coinfected with HIV around the world, not only in injecting drug users, but also in MSM, and to see if there are phylogenetic distinctions between the two groups. Some studies carried out on this subject have found monophyletic clusters within MSM (Danta *et al.*, 2007; Urbanus *et al.*, 2009), and it has been demonstrated that HCV in MSM is strongly associated with multiple sexual partners and a combination of other high risk behaviours (Danta *et al.*, 2007; Pando *et al.*, 2006; Urbanus *et al.*, 2009), which could also explain the monophyletic clusters. Another report found higher spatial HCV heterogeneity in coinfected injecting drug users, due to higher prevalence in needle sharing practises and other risk behaviours (Tan *et al.*, 2008). More studies need to be carried out to establish a more representative global picture of the epidemiology of HCV in individuals coinfected with HIV. The small number of HCV-positive

samples found in the HIV-infected cohort from 2005-2009, could imply a low prevalence of HCV infection among MSM on the island at the moment. However, as this risk group is getting increasing attention with regards to HCV transmission in Europe, this prevalence estimate could also change in Cyprus in the near future.

Due to the small sampling size of the high-risk cohorts, significant epidemiological conclusions cannot be made. A more accurate and representative depiction of the genotype distribution and sequence diversity in these groups would be obtained from larger sampling sets and the availability of more sequences in the public databases for comparison. Given that the countries of origin of many participants in this study are in Eastern Europe, it would be useful to compare their strains to variants retrieved within these countries, but such publicly available data is lacking. Another limitation to detailed analysis is the fact that epidemiological data is based on what the participants state in their questionnaires. Many individuals do not know how, when or where they were infected with HCV, and are sometimes not aware enough to make an educated guess. Also, much of the information given may be lacking in detail or truth, e. g. participants who avoid stating that they have practised injecting drug abuse or certain unsafe sexual behaviours to their doctor or interviewer, due to the social stigma associated with these behaviours. This means that valuable information required for a full epidemiological analysis is missing.

Overall, the results from this part of the study demonstrate a broad genotype distribution (Table 8) and high genetic heterogeneity of the HCV infection in Cyprus, similar to the pattern seen for HIV (Kousiappa *et al.*, 2009a). Samples were taken from epidemiologically diverse populations, and genotypes 1a, 1b and 3a appear to circulate in individuals regardless of their country of origin. The phylogenetic analyses illustrate multiple points of introduction of HCV into the Cypriot population. The existence of small monophyletic clusters demonstrates multiple exposures to needle or syringe sharing between drug users, highlighting the risk of widespread transmission in this high-risk cohort. The overall sequence heterogeneity in HCV strains found in Cyprus is probably due to imported strains fromimmigrants, repatriated Cypriots, Cypriots travelling abroad and the large tourism industry. Also, the diverse ethnic background of the study group and the finding of the possible 2k/1b recombinant strains emphasise the impact of immigration from Eastern Europe and increasing occurrence of injecting drug abuse in Cyprus on the island's HCV situation. Further investigation would be necessary with more extended sampling groups to obtain a more complete picture of the epidemiology of HCV among various cohorts on the island and

rule out the presence of more genotypes and subtypes in these populations. The results here support that the point of entry of HCV on the island or in any of the cohorts investigated was not due to a single-transmission introduction.

Table 8. Genotype distribution among cohorts

Subtype	No. of cases
General population	
1a	13
1b	39
1	1
2c	1
3a	16
4a	6
4	4
5a	1
RF_2k/1b	2
Intravenous drug users	
1b	6
3a	8
Incarcerated population	
1a	2
1b	5
3a	9
HIV/HCV coinfected population	
1a	3
1b	2
3a	2

Also found in the investigation of partial sequences from samples of the general population was evidence for certain uncommon strains: putative 2k/1b recombinants, an unclassified genotype-1-like strain, and unclassified genotype 4 variants. These samples were selected to be characterised in more detail by near-full genome sequencing and analysis, described in the second section of the thesis. This was done by designing strain-specific RT nested PCR and sequencing protocols, and variants from all three categories of HCV type

were amplified and sequenced. The sequences were analysed for recombination events and for phylogenetic relationships with sequences from the public databases.

The sequences of the putative recombinants and epidemiological characteristics of the patients they were isolated from were similar to those of strains of the known 2k/1b recombinant found to be circulating among Eastern European injecting drug users. The strain was originally detected in St. Petersburg (Kalinina *et al.*, 2002), and later sequenced along the full length of its genome (Kalinina *et al.*, 2004), and from all HCV recombinants found to date, it is the only one that has since been detected in numerous infections in various countries. It appears to be circulating among individuals from Eastern Europe and countries of the former Soviet Union, and its associated risk behaviour is intravenous drug use (Kurbanov *et al.*, 2008a; Kurbanov *et al.*, 2008b; Moreau *et al.*, 2006; Morel *et al.*, 2010; Tallo *et al.*, 2007). The discovery of these strains in Cyprus emphasises the increasing transmission of this type within Europe and its association with this risk behaviour. Also, as only two genome sequences have been submitted to GenBank to date, the sequences described here will double the near-full genome sequence information available for this strain in public databases.

With just two full 2k/1b genomes available until now (Kalinina *et al.*, 2004; Morel *et al.*, 2010), it has been very difficult to draw any conclusions regarding its molecular evolution, but with the addition of two more near-full genomes from this study, hypotheses about this can begin to be formulated. Phylogenetic analysis performed on the available 2k/1b sequences of near-full genome or partial sequences, showed that the recombinant strains demonstrated a low evolutionary divergence. Also, the strains were tightly grouped together, forming distinct subclusters within the respective parental subtype groups, in concordance with another study that investigated the 2k/1b molecular epidemiology (Kurbanov *et al.*, 2008b). However, the strains here also demonstrated statistically significant biphyletic clustering. This was observed for the near-full genome sequences and the 1b sequences on the 3′ side of the recombination point, but not for the 2k sequences on the 5′ side. Bearing in mind that not all 2k/1b strains found have been sequenced fully, the data available for these analyses was not the same, and was also limited. Only in the analysis for the sequences 5′ of the recombination site were there more strains available than the ones that have been fully sequenced. Therefore, from the data available including the strains discovered in this study, the observed difference in clustering behaviour between the regions on either side of the recombination point could possibly signify that the original recombination event took place between the 2k quasispecies and genetically diverse 1b genomes, perhaps due to multiple points of infection, and the spread from two of

these recombination events is represented in the strains that have been sequenced to date. The significance of the two strains discovered in Cyprus is the provision of more sequence information regarding the 2k/1b recombinant strain, and providing a different view on the evolution of this strain than that which has been held to date. With the addition of more near-full genomes within this recombinant type, the evolutionary picture of this strain will become more apparent.

This particular strain is comprised of 1b sequence in its NS5A region, where the ISDR lies. However, the response of this, or of any, recombinant to anti-HCV treatment is not known, and only two studies mention the therapeutic response of 2k/1b (Kurbanov *et al.*, 2008b; Morel *et al.*, 2010), resulting in conflicting conclusions. The increasing report of 2k/1b infections in a high-risk population, stress the fact that HCV recombinant strains should be given greater consideration in their detection, diagnosis, clinical management and epidemiology (Kalinina *et al.*, 2002; Kurbanov *et al.*, 2008b; Sentandreu *et al.*, 2008).

Furthermore, recombinant forms of HCV are not included in the considerations for the criteria of HCV classification and nomenclature (Kuiken and Simmonds, 2009). The use of the corresponding HIV system for naming recombinant strains has been suggested, making the strain in question RF1_1b2k, but a set of standardised rules needs to be in place as increasing cases of recombination are being reported and included in phylogenetic analyses.

One strain in the general population group was found to belong in the genotype 1 group, but did not classify with any known subtypes. The route of HCV transmission in this case was not known by the patient, and so no particular risk behaviour or geographic origin can be associated with it. Also, considering that it has only ever been found in one case of infection to date, it is not yet significant from an epidemiological point of view, as it may not be spreading in the population. If any new cases are discovered, however, this may raise an alarm due to implications it may have on response to therapy. Near-full genome sequence analysis revealed no evidence of recombination. Phylogenetic analysis showed that this isolate groups closely with genotype 1 with a high bootstrap support, but more genetically distant from the rest of genotype 1 subtypes than they are from each other. The percentage genetic variability from genotype 1 (~31%) implies that this strain is on the verge of belonging to a new HCV genotype (Simmonds *et al.*, 2005). This unclassified strain is more likely a subtype of genotype 1 which has not been discovered until now. More epidemiologically distinct examples of this strain are required for confirmed designation. Non-a, non-b subtypes of genotype 1 found to date are very few and the lack of full or near-full genome sequence data

publicly available for them presents an issue in classifying the variant described here. Novel subtypes of genotype 1 are not commonly found, and so the discovery of this strain stresses the continuous evolution of this virus and the near-full sequence data is significant for evolutionary investigations.

A cluster of unclassified variants belonging to genotype 4 were also identified and two of these were sequenced along the near-full length genome. These sequences are distinct from any known genotype 4 subtypes and are considered to belong to a novel genotype 4 subtype. HCV genotype 4 is prevalent in the Middle East and Africa, where it exhibits great genetic diversity as it has been endemic in these regions for a long time, and has recently spread to Europe and America due to immigration and intravenous drug use (Ansaldi *et al.*, 2005; Fernandez-Arcas *et al.*, 2006; Franco *et al.*, 2007; Kamal and Nasser, 2008; Katsoulidou *et al.*, 2006; Lyra *et al.*, 2004; Murphy *et al.*, 2007b; Nicot *et al.*, 2005; Payan *et al.*, 2005; Savvas *et al.*, 2005). This spread of genotype 4 strains is also seen in this study. The sequence analysis of the strains in question revealed no recombination events and a close genetic relationship to subtypes 4q and 4l. Like these, the novel genotype 4 variant probably evolved from a common ancestor originating in Africa. The genetic variation between these subtypes and the strains found here is smaller than the criteria for subtype discrimination (Simmonds *et al.*, 2005), but they belong to distinct subtypes, representing the epidemiological associations of these variants and possibly indicating a continual genetic variation of HCV with many variants still missing (de Bruijne *et al.*, 2009b; Li *et al.*, 2009a). Based on the results of the sequence analysis, and the current picture of subtype classification in HCV genotype 4, it is proposed that this group of strains should be designated as new subtype 4v. Epidemiologically distinct infections of this variant have been found in Canada and Britain, and here we present two near-full genome reference sequences of this novel type for the first time. The sequences presented here can be used as references for HCV classification, but also for diagnostic and clinical purposes. As more HCV variants are identified, new data will be obtained and can be used to develop additional nomenclature and classification regulations.

Different levels of within-group diversity are found between genotypes, and different relationships between them. Specifying a degree of sequence divergence below which a subtype designation is made, and above which a new genotype is assigned, is difficult and arbitrary (Kuiken and Simmonds, 2009). Uncommon strains of HCV have been found in Cyprus, and near-full genome sequence analysis revealed that they represent both contributors to HCV genetic variation, i.e. recombination events and a high mutation rate. Two strains

belong to the increasingly circulating 2k/1b form of recombinant, and the sequences presented here make a significant contribution to the sequence data available on this variant, allowing more sophisticated evolutionary hypotheses to be generated regarding its origins. A novel genotype-1-like strain was identified here for the first time, unrelated to any other strain on public databases. Additional isolates of a previously discovered unclassified genotype 4 variant have also been identified and near-full reference sequences are presented here for the first time, leading to their possible classification as a novel subtype, 4v. These findings confirm that although the genetic variability of HCV is known to be high, and numerous subtypes have been named to date, novel subtypes still continue to be uncovered. These analyses add to our knowledge of global HCV genetic diversity, which is vital for optimisation of treatment regimes.

In conclusion, studies on the hepatitis C virus are important, as it is a globally prevalent disease, with high morbidity and mortality associated with it. As this is expected to rise in the near future, this disease will continue to cause a major problem for public health. The results of this thesis reveal the molecular epidemiology of the HCV infection in Cyprus and demonstrate its continual evolution and widespread transmission. The situation of the HCV infection in Cyprus highlights the changing epidemiology of HCV, with genetically variant strains finding their way into Europe from the Middle East, Africa and Eastern Europe, by means of injecting drug abuse and immigration. Also, the broad genotype distribution and multiple points of infection have implications on the clinical management of the infection, and highlight the urgent need for awareness, risk reduction among high-risk populations, and prevention strategies. Following publication and submission to public databases, the results presented here will add to the knowledge of HCV sequence variation, which is significant for classification and nomenclature purposes.

Future directions of this research could include a continuation to study the change of the HCV molecular epidemiology in Cyprus over time, including prevalence studies. As new drugs are made available, longitudinal samples from patients on therapy could determine genetic variations of the virus caused by selective pressure, or the development of genotypic drug resistance in variants of different genotypes. Efforts to generate full genome sequences from HCV genotypes and subtypes that are not prevalent in the public databases will surely significantly contribute to our understanding of the sequence diversity of HCV and for the rational design of vaccines and novel therapies.

Bibliography

Abdel-Hamid, M., El-Daly, M., Molnegren, V., El-Kafrawy, S., Abdel-Latif, S., Esmat, G., Strickland, G. T., Loffredo, C., Albert, J., and Widell, A. (2007). Genetic diversity in hepatitis C virus in Egypt and possible association with hepatocellular carcinoma. *J Gen Virol* **88**(Pt 5)**,** 1526-31.

Aceijas, C., and Rhodes, T. (2007). Global estimates of prevalence of HCV infection among injecting drug users. *Int J Drug Policy* **18**(5)**,** 352-8.

Agnello, V., Abel, G., Elfahal, M., Knight, G. B., and Zhang, Q. X. (1999). Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci U S A* **96**(22)**,** 12766-71.

Aitken, C. K., Lewis, J., Tracy, S. L., Spelman, T., Bowden, D. S., Bharadwaj, M., Drummer, H., and Hellard, M. (2008). High incidence of hepatitis C virus reinfection in a cohort of injecting drug users. *Hepatology* **48**(6)**,** 1746-52.

Alter, H. (1999). Discovery of non-A, non-B hepatitis and identification of its etiology. *Am J Med* **107**(6B)**,** 16S-20S.

Alter, M. J. (1997). The epidemiology of acute and chronic hepatitis C. *Clin Liver Dis* **1**(3)**,** 559-68, vi-vii.

Alter, M. J. (2002). Prevention of spread of hepatitis C. *Hepatology* **36**(5 Suppl 1)**,** S93-8.

Alter, M. J. (2007). Epidemiology of hepatitis C virus infection. *World J Gastroenterol* **13**(17)**,** 2436-41.

Altindis, M., Yilmaz, S., Dikengil, T., Acemoglu, H., and Hosoglu, S. (2006). Seroprevalence and genotyping of hepatitis B, hepatitis C and HIV among healthy population and Turkish soldiers in Northern Cyprus. *World J Gastroenterol* **12**(42)**,** 6792-6.

Altuglu, I., Soyler, I., Ozacar, T., and Erensoy, S. (2008). Distribution of hepatitis C virus genotypes in patients with chronic hepatitis C infection in Western Turkey. *Int J Infect Dis* **12**(3)**,** 239-44.

Ansaldi, F., Bruzzone, B., Salmaso, S., Rota, M. C., Durando, P., Gasparini, R., and Icardi, G. (2005). Different seroprevalence and molecular epidemiology patterns of hepatitis C virus infection in Italy. *J Med Virol* **76**(3)**,** 327-32.

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

Babudieri, S., Longo, B., Sarmati, L., Starnini, G., Dori, L., Suligoi, B., Carbonara, S., Monarca, R., Quercia, G., Florenzano, G., Novati, S., Sardu, A., Iovinella, V., Casti, A., Romano, A., Uccella, I., Maida, I., Brunetti, B., Mura, M. S., Andreoni, M., and Rezza, G. (2005). Correlates of HIV, HBV, and HCV infections in a prison inmate population: results from a multicentre study in Italy. *J Med Virol* **76**(3)**,** 311-7.

Bartenschlager, R., Frese, M., and Pietschmann, T. (2004). Novel insights into hepatitis C virus replication and persistence. *Adv Virus Res* **63,** 71-180.

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

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

Bartosch, B., Dubuisson, J., and Cosset, F. L. (2003). Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *J Exp Med* **197**(5)**,** 633-42.

Bassett, S. E., Guerra, B., Brasky, K., Miskovsky, E., Houghton, M., Klimpel, G. R., and Lanford, R. E. (2001). Protective immune response to hepatitis C virus in chimpanzees rechallenged following clearance of primary infection. *Hepatology* **33**(6)**,** 1479-87.

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

Bukh, J., Miller, R. H., and Purcell, R. H. (1995). Biology and genetic heterogeneity of hepatitis C virus. *Clin Exp Rheumatol* **13**(Suppl 13)**,** S3-7.

Cenci, M., Massi, M., Alderisio, M., De Soccio, G., and Recchia, O. (2007). Prevalence of hepatitis C virus (HCV) genotypes and increase of type 4 in central Italy: an update and report of a new method of HCV genotyping. *Anticancer Res* **27**(2)**,** 1219-22.

Chamberlain, R. W., Adams, N., Saeed, A. A., Simmonds, P., and Elliott, R. M. (1997a). Complete nucleotide sequence of a type 4 hepatitis C virus variant, the predominant genotype in the Middle East. *J Gen Virol* **78**(Pt 6)**,** 1341-7.

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

Chang, M. H. (1996). Mother-to-infant transmission of hepatitis C virus. *Clin Invest Med* **19**(5)**,** 368-72.

Charlton, M. (2001). Hepatitis C infection in liver transplantation. *Am J Transplant* **1**(3)**,** 197-203.

Chen, S. L., and Morgan, T. R. (2006). The natural history of hepatitis C virus (HCV) infection. *Int J Med Sci* **3**(2)**,** 47-52. Epub 2006 Apr 1.

Chevaliez, S., and Pawlotsky, J. M. (2006). HCV Genome and Life Cycle. *In* "Hepatitis C Viruses: Genomes and Molecular Biology" (S. L. Tan, Ed.). Horizon Bioscience, Norfolk (UK).

Chlabicz, S., Flisiak, R., Kowalczuk, O., Wiercinska-Drapalo, A., Pytel-Krolczuk, B., Prokopowicz, D., and Chyczewski, L. (2008). High prevalence of genotype 4 among hepatitis C virus-infected intravenous drug users in north-eastern Poland. *J Med Virol* **80**(4)**,** 615-8.

Choi, J., Xu, Z., and Ou, J. H. (2003). Triple decoding of hepatitis C virus RNA by programmed translational frameshifting. *Mol Cell Biol* **23**(5)**,** 1489-97.

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

Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P. J., and et al. (1991). Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A* **88**(6)**,** 2451-5.

Cochrane, A., Searle, B., Hardie, A., Robertson, R., Delahooke, T., Cameron, S., Tedder, R. S., Dusheiko, G. M., De Lamballerie, X., and Simmonds, P. (2002). A genetic analysis of hepatitis C virus transmission between injection drug users. *J Infect Dis* **186**(9)**,** 1212-21.

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

Cristina, J., and Colina, R. (2006). Evidence of structural genomic region recombination in Hepatitis C virus. *Virol J* **3,** 53.

Crofts, N., Aitken, C. K., and Kaldor, J. M. (1999). The force of numbers: why hepatitis C is spreading among Australian injecting drug users while HIV is not. *Med J Aust* **170**(5)**,** 220-1.

Danta, M., Brown, D., Bhagani, S., Pybus, O. G., Sabin, C. A., Nelson, M., Fisher, M., Johnson, A. M., and Dusheiko, G. M. (2007). Recent epidemic of acute hepatitis C virus in HIV-positive men who have sex with men linked to high-risk sexual behaviours. *AIDS* **21**(8)**,** 983-91.

de Bruijne, J., Schinkel, J., Prins, M., Koekkoek, S. M., Aronson, S. J., van Ballegooijen, M. W., Reesink, H. W., Molenkamp, R., and van de Laar, T. J. (2009b). Emergence of hepatitis C virus genotype 4: phylogenetic analysis reveals three distinct epidemiological profiles. *J Clin Microbiol*.

de Bruijne, J., Weegink, C. J., Jansen, P. L., and Reesink, H. W. (2009a). New developments in the antiviral treatment of hepatitis C. *Vox Sang* **97**(1)**,** 1-12.

De Francesco, R., and Carfi, A. (2007). Advances in the development of new therapeutic agents targeting the NS3-4A serine protease or the NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *Adv Drug Deliv Rev* **59**(12)**,** 1242-62.

De Francesco, R., Tomei, L., Altamura, S., Summa, V., and Migliaccio, G. (2003). Approaching a new era for hepatitis C virus therapy: inhibitors of the NS3-4A serine protease and the NS5B RNA-dependent RNA polymerase. *Antiviral Res* **58**(1)**,** 1-16.

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E. J., Wensing, A. M., van de Vijver, D. A., Boucher, C. A., Camacho, R., and Vandamme, A. M. (2005). An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* **21**(19)**,** 3797-800.

Deuffic-Burban, S., Mohamed, M. K., Larouze, B., Carrat, F., and Valleron, A. J. (2006). Expected increase in hepatitis C-related mortality in Egypt due to pre-2000 infections. *J Hepatol* **44**(3)**,** 455-61.

Deuffic-Burban, S., Poynard, T., Sulkowski, M. S., and Wong, J. B. (2007). Estimating the future health burden of chronic hepatitis C and human immunodeficiency virus infections in the United States. *J Viral Hepat* **14**(2)**,** 107-15.

Di Martino, V., Rufat, P., Boyer, N., Renard, P., Degos, F., Martinot-Peignoux, M., Matheron, S., Le Moing, V., Vachon, F., Degott, C., Valla, D., and Marcellin, P. (2001). The influence of human immunodeficiency virus coinfection on chronic hepatitis C in injection drug users: a long-term retrospective cohort study. *Hepatology* **34**(6)**,** 1193-9.

Dowd, K. A., Hershow, R. C., Yawetz, S., Larussa, P., Diaz, C., Landesman, S. H., Paul, M. E., Read, J. S., Lu, M., Thomas, D. L., Netski, D. M., and Ray, S. C. (2008). Maternal neutralizing antibody and transmission of hepatitis C virus to infants. *J Infect Dis* **198**(11)**,** 1651-5.

Dubuisson, J. (2007). Hepatitis C virus proteins. *World J Gastroenterol* **13**(17)**,** 2406-15.

Dubuisson, J., Helle, F., and Cocquerel, L. (2008). Early steps of the hepatitis C virus life cycle. *Cell Microbiol* **10**(4)**,** 821-7.

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

England, K., Thorne, C., and Newell, M. L. (2006). Vertically acquired paediatric coinfection with HIV and hepatitis C virus. *Lancet Infect Dis* **6**(2)**,** 83-90.

Esteban, J. I., Sauleda, S., and Quer, J. (2008). The changing epidemiology of hepatitis C virus infection in Europe. *J Hepatol* **48**(1)**,** 148-62.

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

Farci, P. (2002). 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;244:359-362]. *J Hepatol* **36**(5)**,** 582-5.

Farci, P., Alter, H. J., Wong, D., Miller, R. H., Shih, J. W., Jett, B., and Purcell, R. H. (1991). A long-term study of hepatitis C virus replication in non-A, non-B hepatitis. *N Engl J Med* **325**(2)**,** 98-104.

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

Feld, J. J., and Hoofnagle, J. H. (2005). Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* **436**(7053)**,** 967-72.

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

Ferenci, P. (2004). Predictors of response to therapy for chronic hepatitis C. *Semin Liver Dis* **24**(Suppl 2)**,** 25-31.

Fernandez-Arcas, N., Lopez-Siles, J., Trapero, S., Ferraro, A., Ibanez, A., Orihuela, F., Maldonado, J., and Alonso, A. (2006). High prevalence of hepatitis C virus subtypes 4c and 4d in Malaga (Spain): phylogenetic and epidemiological analyses. *J Med Virol* **78**(11)**,** 1429-35.

Ferrero, S., Lungaro, P., Bruzzone, B. M., Gotta, C., Bentivoglio, G., and Ragni, N. (2003). Prospective study of mother-to-infant transmission of hepatitis C virus: a 10-year survey (1990-2000). *Acta Obstet Gynecol Scand* **82**(3)**,** 229-34.

Fleckenstein, J. (2004). Chronic hepatitis C in African Americans and other minority groups. *Curr Gastroenterol Rep* **6**(1)**,** 66-70.

Franco, S., Tural, C., Clotet, B., and Martinez, M. A. (2007). Complete nucleotide sequence of genotype 4 hepatitis C viruses isolated from patients co-infected with human immunodeficiency virus type 1. *Virus Res* **123**(2)**,** 161-9.

Frank, C., Mohamed, M. K., Strickland, G. T., Lavanchy, D., Arthur, R. R., Magder, L. S., El Khoby, T., Abdel-Wahab, Y., Aly Ohn, E. S., Anwar, W., and Sallam, I. (2000). The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* **355**(9207)**,** 887-91.

Fried, M. W., Shiffman, M. L., Reddy, K. R., Smith, C., Marinos, G., Goncales, F. L., Jr., Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J., and Yu, J. (2002). Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* **347**(13)**,** 975-82.

Gardner, J. P., Durso, R. J., Arrigale, R. R., Donovan, G. P., Maddon, P. J., Dragic, T., and Olson, W. C. (2003). L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. *Proc Natl Acad Sci U S A* **100**(8)**,** 4498-503.

Genovese, D., Dettori, S., Argentini, C., Villano, U., Chionne, P., Angelico, M., and Rapicetta, M. (2005). Molecular epidemiology of hepatitis C virus genotype 4 isolates in Egypt and analysis of the variability of envelope proteins E1 and E2 in patients with chronic hepatitis. *J Clin Microbiol* **43**(4)**,** 1902-9.

Ghosn, J., Thibault, V., Delaugerre, C., Fontaine, H., Lortholary, O., Rouzioux, C., Pol, S., and Chaix, M. L. (2008). Sexually transmitted hepatitis C virus superinfection in HIV/hepatitis C virus co-infected men who have sex with men. *AIDS* **22**(5)**,** 658-61.

Hagan, H., Des Jarlais, D. C., Stern, R., Lelutiu-Weinberger, C., Scheinmann, R., Strauss, S., and Flom, P. L. (2007). HCV synthesis project: preliminary analyses of HCV prevalence in relation to age and duration of injection. *Int J Drug Policy* **18**(5)**,** 341-51.

Hao, W., and Duggal, R. (2009). High-throughput screening of HCV RNA replication inhibitors by means of a reporter replicon system. *Methods Mol Biol* **510,** 243-50.

Harrison, S. A. (2007). Small molecule and novel treatments for chronic hepatitis C virus infection. *Am J Gastroenterol* **102**(10)**,** 2332-8.

Health Protection Agency, U. K. (2008). Hepatitis C in England: an update. Annual Report.

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

Hinrichsen, H., Benhamou, Y., Wedemeyer, H., Reiser, M., Sentjens, R. E., Calleja, J. L., Forns, X., Erhardt, A., Cronlein, J., Chaves, R. L., Yong, C. L., Nehmiz, G., and Steinmann, G. G. (2004). Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients. *Gastroenterology* **127**(5)**,** 1347-55.

Hmaied, F., Legrand-Abravanel, F., Nicot, F., Garrigues, N., Chapuy-Regaud, S., Dubois, M., Njouom, R., Izopet, J., and Pasquier, C. (2007). Full-length genome sequences of hepatitis C virus subtype 4f. *J Gen Virol* **88**(Pt 11)**,** 2985-90.

Holsen, D. S., Harthug, S., and Myrmel, H. (1993). Prevalence of antibodies to hepatitis C virus and association with intravenous drug abuse and tattooing in a national prison in Norway. *Eur J Clin Microbiol Infect Dis* **12**(9)**,** 673-6.

Houghton, M. (2009). The long and winding road leading to the identification of the hepatitis C virus. *J Hepatol* **51**(5)**,** 939-48.

Houghton, M., and Abrignani, S. (2005). Prospects for a vaccine against the hepatitis C virus. *Nature* **436**(7053)**,** 961-6.

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8)**,** 754-5.

Huy, T. T., and Abe, K. (2004). Molecular epidemiology of hepatitis B and C virus infections in Asia. *Pediatr Int* **46**(2)**,** 223-30.

Idrees, M. (2008). Development of an improved genotyping assay for the detection of hepatitis C virus genotypes and subtypes in Pakistan. *J Virol Methods* **150**(1-2)**,** 50-6.

Indolfi, G., and Resti, M. (2009). Perinatal transmission of hepatitis C virus infection. *J Med Virol* **81**(5)**,** 836-43.

Judd, A., Hutchinson, S., Wadd, S., Hickman, M., Taylor, A., Jones, S., Parry, J. V., Cameron, S., Rhodes, T., Ahmed, S., Bird, S., Fox, R., Renton, A., Stimson, G. V., and Goldberg, D. (2005). Prevalence of, and risk factors for, hepatitis C virus infection among recent initiates to injecting in London and Glasgow: cross sectional analysis. *J Viral Hepat* **12**(6)**,** 655-62.

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

Kalinina, O., Norder, H., and Magnius, L. O. (2004). Full-length open reading frame of a recombinant hepatitis C virus strain from St Petersburg: proposed mechanism for its formation. *J Gen Virol* **85**(Pt 7)**,** 1853-7.

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

Kalinina, O., Norder, H., Vetrov, T., Zhdanov, K., Barzunova, M., Plotnikova, V., Mukomolov, S., and Magnius, L. O. (2001). Shift in predominating subtype of HCV from 1b to 3a in St. Petersburg mediated by increase in injecting drug use. *J Med Virol* **65**(3)**,** 517-24.

Kamal, S. M., and Nasser, I. A. (2008). Hepatitis C genotype 4: What we know and what we don't yet know. *Hepatology*.

Kapadia, S. B., and Chisari, F. V. (2005). Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. *Proc Natl Acad Sci U S A* **102**(7)**,** 2561-6.

Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K. (1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci U S A* **87**(24)**,** 9524-8.

Katsoulidou, A., Sypsa, V., Tassopoulos, N. C., Boletis, J., Karafoulidou, A., Ketikoglou, I., Tsantoulas, D., Vafiadi, I., Hatzis, G., Skoutelis, A., Akriviadis, E., Vasiliadis, T., Kitis, G., Magiorkinis, G., and Hatzakis, A. (2006). Molecular epidemiology of hepatitis C virus (HCV) in Greece: temporal trends in HCV genotype-specific incidence and molecular characterization of genotype 4 isolates. *J Viral Hepat* **13**(1)**,** 19-27.

Katze, M. G., He, Y., and Gale, M., Jr. (2002). Viruses and interferon: a fight for supremacy. *Nat Rev Immunol* **2**(9)**,** 675-87.

Kim, A. Y., and Chung, R. T. (2009). Coinfection with HIV-1 and HCV - a one-two punch. *Gastroenterology* **137**(3)**,** 795-814.

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

Koletzki, D., Dumont, S., Vermeiren, H., Peixe, P., Nina, J., Camacho, R. J., and Stuyver, L. J. (2009). Full genome sequence of three isolates of hepatitis C virus subtype 4b from Portugal. *Arch Virol* **154**(1)**,** 127-32.

Kousiappa, I., van de Vijver, D. A., Demetriades, I., and Kostrikis, L. G. (2009a). Genetic analysis of HIV type 1 strains from newly infected untreated patients in cyprus: high genetic diversity and low prevalence of drug resistance. *AIDS Res Hum Retroviruses* **25**(1)**,** 23-35.

Kousiappa, I., Van De Vijver, D. A., and Kostrikis, L. G. (2009b). Near full-length genetic analysis of HIV sequences derived from Cyprus: evidence of a highly polyphyletic and evolving infection. *AIDS Res Hum Retroviruses* **25**(8)**,** 727-40.

Koziel, M. J., and Peters, M. G. (2007). Viral hepatitis in HIV infection. *N Engl J Med* **356**(14)**,** 1445-54.

Kretzschmar, M., and Wiessing, L. (2008). New challenges for mathematical and statistical modeling of HIV and hepatitis C virus in injecting drug users. *AIDS* **22**(13)**,** 1527-37.

Kuiken, C., and Simmonds, P. (2009). Nomenclature and numbering of the hepatitis C virus. *Methods Mol Biol* **510,** 33-53.

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

Kuntzen, T., Berical, A., Ndjomou, J., Bennett, P., Schneidewind, A., Lennon, N., Birren, B. W., Kuiken, C., Henn, M. R., Simmonds, P., and Allen, T. M. (2008b). A set of reference sequences for the hepatitis C genotypes 4d, 4f, and 4k covering the full open reading frame. *J Med Virol* **80**(8)**,** 1370-8.

Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A. M., Young, S. K., Lee, B., Heckerman, D., Carlson, J., Reyor, L. L., Kleyman, M., McMahon, C. M., Birch, C., Schulze Zur Wiesch, J., Ledlie, T., Koehrsen, M., Kodira, C., Roberts, A. D., Lauer, G.

M., Rosen, H. R., Bihl, F., Cerny, A., Spengler, U., Liu, Z., Kim, A. Y., Xing, Y., Schneidewind, A., Madey, M. A., Fleckenstein, J. F., Park, V. M., Galagan, J. E., Nusbaum, C., Walker, B. D., Lake-Bakaar, G. V., Daar, E. S., Jacobson, I. M., Gomperts, E. D., Edlin, B. R., Donfield, S. M., Chung, R. T., Talal, A. H., Marion, T., Birren, B. W., Henn, M. R., and Allen, T. M. (2008a). Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. *Hepatology* **48**(6)**,** 1769-78.

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

Kurbanov, F., Tanaka, Y., Chub, E., Maruyama, I., Azlarova, A., Kamitsukasa, H., Ohno, T., Bonetto, S., Moreau, I., Fanning, L. J., Legrand-Abravanel, F., Izopet, J., Naoumov, N., Shimada, T., Netesov, S., and Mizokami, M. (2008b). Molecular epidemiology and interferon susceptibility of the natural recombinant hepatitis C virus strain RF1_2k/1b. *J Infect Dis* **198**(10)**,** 1448-56.

Kurbanov, F., Tanaka, Y., Sugauchi, F., Kato, H., Ruzibakiev, R., Zalyalieva, M., Yunusova, Z., and Mizokami, M. (2003). Hepatitis C virus molecular epidemiology in Uzbekistan. *J Med Virol* **69**(3)**,** 367-75.

Lai, M. E., Mazzoleni, A. P., Argiolu, F., De Virgilis, S., Balestrieri, A., Purcell, R. H., Cao, A., and Farci, P. (1994). Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. *Lancet* **343**(8894)**,** 388-90.

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

Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., and Leigh Brown, A. J. (2008). Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* **5**(3)**,** e50.

Li, C., Fu, Y., Lu, L., Ji, W., Yu, J., Hagedorn, C. H., and Zhang, L. (2006). Complete genomic sequences for hepatitis C virus subtypes 6e and 6g isolated from Chinese patients with injection drug use and HIV-1 co-infection. *J Med Virol* **78**(8)**,** 1061-9.

Li, C., Lu, L., Wu, X., Wang, C., Bennett, P., Lu, T., and Murphy, D. (2009a). Complete genomic sequences for hepatitis C virus subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t. *J Gen Virol* **90**(Pt 8)**,** 1820-6.

Li, C., Lu, L., Zhang, X., and Murphy, D. (2009b). Entire genome sequences of two new HCV subtypes, 6r and 6s, and characterization of unique HVR1 variation patterns within genotype 6. *J Viral Hepat* **16**(6)**,** 406-17.

Lindenbach, B. D., Evans, M. J., Syder, A. J., Wolk, B., Tellinghuisen, T. L., Liu, C. C., Maruyama, T., Hynes, R. O., Burton, D. R., McKeating, J. A., and Rice, C. M. (2005). Complete replication of hepatitis C virus in cell culture. *Science* **309**(5734)**,** 623-6.

Lindenbach, B. D., and Rice, C. M. (2005). Unravelling hepatitis C virus replication from genome to function. *Nature* **436**(7053)**,** 933-8.

Liu, J. Y., Lin, H. H., Liu, Y. C., Lee, S. S., Chen, Y. L., Hung, C. C., Ko, W. C., Huang, C. K., Lai, C. H., Chen, Y. S., Shih, Y. L., Chung, H. C., Liang, S. H., and Lin, J. N. (2008). Extremely high prevalence and genetic diversity of hepatitis C virus infection among HIV-infected injection drug users in Taiwan. *Clin Infect Dis* **46**(11)**,** 1761-8.

Lozach, P. Y., Amara, A., Bartosch, B., Virelizier, J. L., Arenzana-Seisdedos, F., Cosset, F. L., and Altmeyer, R. (2004). C-type lectins L-SIGN and DC-SIGN capture and

transmit infectious hepatitis C virus pseudotype particles. *J Biol Chem* **279**(31)**,** 32035-45.

Lozach, P. Y., Lortat-Jacob, H., de Lacroix de Lavalette, A., Staropoli, I., Foung, S., Amara, A., Houles, C., Fieschi, F., Schwartz, O., Virelizier, J. L., Arenzana-Seisdedos, F., and Altmeyer, R. (2003). DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. *J Biol Chem* **278**(22)**,** 20358-66.

Lyra, A. C., Ramrakhiani, S., Bacon, B. R., and Di Bisceglie, A. M. (2004). Infection with hepatitis C virus genotype 4 in the United States. *J Clin Gastroenterol* **38**(1)**,** 68-71.

Manns, M. P., Foster, G. R., Rockstroh, J. K., Zeuzem, S., Zoulim, F., and Houghton, M. (2007). The way forward in HCV treatment - finding the right path. *Nat Rev Drug Discov* **6**(12)**,** 991-1000.

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

Mathei, C., Wollants, E., Verbeeck, J., Van Ranst, M., Robaeys, G., Van Damme, P., and Buntinx, F. (2005). Molecular epidemiology of hepatitis C among drug users in Flanders, Belgium: association of genotype with clinical parameters and with sex- and drug-related risk behaviours. *Eur J Clin Microbiol Infect Dis* **24**(8)**,** 514-22.

Mboto, C. I., Davies, A., Fielder, M., and Jewell, A. P. (2006). Human immunodeficiency virus and hepatitis C co-infection in sub-Saharan West Africa. *Br J Biomed Sci* **63**(1)**,** 29-37.

McCaw, R., Moaven, L., Locarnini, S. A., and Bowden, D. S. (1997). Hepatitis C virus genotypes in Australia. *J Viral Hepat* **4**(5)**,** 351-7.

McGovern, B. H., Wurcel, A., Kim, A. Y., Schulze zur Wiesch, J., Bica, I., Zaman, M. T., Timm, J., Walker, B. D., and Lauer, G. M. (2006). Acute hepatitis C virus infection in incarcerated injection drug users. *Clin Infect Dis* **42**(12)**,** 1663-70.

McKeating, J. A., Zhang, L. Q., Logvinoff, C., Flint, M., Zhang, J., Yu, J., Butera, D., Ho, D. D., Dustin, L. B., Rice, C. M., and Balfe, P. (2004). Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. *J Virol* **78**(16)**,** 8496-505.

Meier, V., and Ramadori, G. (2009). Hepatitis C virus virology and new treatment targets. *Expert Rev Anti Infect Ther* **7**(3)**,** 329-50.

Memon, M. I., and Memon, M. A. (2002). Hepatitis C: an epidemiological review. *J Viral Hepat* **9**(2)**,** 84-100.

Mercer, D. F., Schiller, D. E., Elliott, J. F., Douglas, D. N., Hao, C., Rinfret, A., Addison, W. R., Fischer, K. P., Churchill, T. A., Lakey, J. R., Tyrrell, D. L., and Kneteman, N. M. (2001). Hepatitis C virus replication in mice with chimeric human livers. *Nat Med* **7**(8)**,** 927-33.

Meuleman, P., Libbrecht, L., De Vos, R., de Hemptinne, B., Gevaert, K., Vandekerckhove, J., Roskams, T., and Leroux-Roels, G. (2005). Morphological and biochemical characterization of a human liver in a uPA-SCID mouse chimera. *Hepatology* **41**(4)**,** 847-56.

Micallef, J. M., Macdonald, V., Jauncey, M., Amin, J., Rawlinson, W., van Beek, I., Kaldor, J. M., White, P. A., and Dore, G. J. (2007). High incidence of hepatitis C virus reinfection within a cohort of injecting drug users. *J Viral Hepat* **14**(6)**,** 413-8.

Miller, E. R., Bi, P., and Ryan, P. (2009a). Hepatitis C virus infection in South Australian prisoners: seroprevalence, seroconversion, and risk factors. *Int J Infect Dis* **13**(2)**,** 201-8.

Miller, E. R., Hellard, M. E., Bowden, S., Bharadwaj, M., and Aitken, C. K. (2009b). Markers and risk factors for HCV, HBV and HIV in a network of injecting drug users in Melbourne, Australia. *J Infect* **58**(5)**,** 375-82.

Monazahian, M., Bohme, I., Bonk, S., Koch, A., Scholz, C., Grethe, S., and Thomssen, R. (1999). Low density lipoprotein receptor as a candidate receptor for hepatitis C virus. *J Med Virol* **57**(3)**,** 223-9.

Monazahian, M., Kippenberger, S., Muller, A., Seitz, H., Bohme, I., Grethe, S., and Thomssen, R. (2000). Binding of human lipoproteins (low, very low, high density lipoproteins) to recombinant envelope proteins of hepatitis C virus. *Med Microbiol Immunol* **188**(4)**,** 177-84.

Moradpour, D., Brass, V., Gosert, R., Wolk, B., and Blum, H. E. (2002). Hepatitis C: molecular virology and antiviral targets. *Trends Mol Med* **8**(10)**,** 476-82.

Moradpour, D., Penin, F., and Rice, C. M. (2007). Replication of hepatitis C virus. *Nat Rev Microbiol* **5**(6)**,** 453-63.

Moreau, I., Hegarty, S., Levis, J., Sheehy, P., Crosbie, O., Kenny-Walsh, E., and Fanning, L. J. (2006). Serendipitous identification of natural intergenotypic recombinants of hepatitis C in Ireland. *Virol J* **3,** 95.

Morel, V., Descamps, V., Francois, C., Fournier, C., Brochot, E., Capron, D., Duverlie, G., and Castelain, S. (2010). Emergence of a genomic variant of the recombinant 2k/1b strain during a mixed Hepatitis C infection: a case report. *J Clin Virol* **47**(4)**,** 382-6.

Moreno, M. P., Casane, D., Lopez, L., and Cristina, J. (2006). Evidence of recombination in quasispecies populations of a Hepatitis C Virus patient undergoing anti-viral therapy. *Virol J* **3,** 87.

Moreno, P., Alvarez, M., Lopez, L., Moratorio, G., Casane, D., Castells, M., Castro, S., Cristina, J., and Colina, R. (2009). Evidence of recombination in Hepatitis C Virus populations infecting a hemophiliac patient. *Virol J* **6,** 203.

Morice, Y., Cantaloube, J. F., Beaucourt, S., Barbotte, L., De Gendt, S., Goncales, F. L., Butterworth, L., Cooksley, G., Gish, R. G., Beaugrand, M., Fay, F., Fay, O., Gonzalez, J. E., Martins, R. M., Dhumeaux, D., Vanderborght, B., Stuyver, L., Sablon, E., de Lamballerie, X., and Pawlotsky, J. M. (2006). Molecular epidemiology of hepatitis C virus subtype 3a in injecting drug users. *J Med Virol* **78**(10)**,** 1296-303.

Murphy, D., Chamberland, J., Dandavino, R., and Sablon, E. (2007a). Abstract 827: A new genotype of hepatitis C virus originating from Central Africa. *Hepatology* **46**(Suppl. 1)**,** 623A.

Murphy, D. G., Willems, B., Deschenes, M., Hilzenrat, N., Mousseau, R., and Sabbah, S. (2007b). Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5' untranslated region sequences. *J Clin Microbiol* **45**(4)**,** 1102-12.

Nakano, T., Lu, L., Liu, P., and Pybus, O. G. (2004). Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J Infect Dis* **190**(6)**,** 1098-108. Epub 2004 Aug 10.

Naoumov, N. V. (1999). Hepatitis C virus infection in Eastern Europe. *J Hepatol* **31 Suppl 1,** 84-7.

Ngo-Giang-Huong, N., Jourdain, G., Sirirungsi, W., Decker, L., Khamduang, W., Le Coeur, S., Sirinontakan, S., Somsamai, R., Pagdi, K., Hemvuttiphan, J., McIntosh, K., Barin, F., and Lallemant, M. (2010). Human immunodeficiency virus-hepatitis C virus co-infection in pregnant women and perinatal transmission to infants in Thailand. *Int J Infect Dis* **14**(7)**,** e602-7.

Nguyen, M. H., and Keeffe, E. B. (2005). Prevalence and treatment of hepatitis C virus genotypes 4, 5, and 6. *Clin Gastroenterol Hepatol* **3**(10 Suppl 2)**,** S97-S101.

Nicot, F., Legrand-Abravanel, F., Sandres-Saune, K., Boulestin, A., Dubois, M., Alric, L., Vinel, J. P., Pasquier, C., and Izopet, J. (2005). Heterogeneity of hepatitis C virus genotype 4 strains circulating in south-western France. *J Gen Virol* **86**(Pt 1)**,** 107-14.

Noppornpanth, S., Lien, T. X., Poovorawan, Y., Smits, S. L., Osterhaus, A. D., and Haagmans, B. L. (2006). Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J Virol* **80**(15)**,** 7569-77.

Noppornpanth, S., Poovorawan, Y., Lien, T. X., Smits, S. L., Osterhaus, A. D., and Haagmans, B. L. (2008). Complete genome analysis of hepatitis C virus subtypes 6t and 6u. *J Gen Virol* **89**(Pt 5)**,** 1276-81.

Oliveira, M. L., Bastos, F. I., Telles, P. R., Hacker Mde, A., Oliveira, S. A., Miguel, J. C., and Yoshida, C. F. (2009). Epidemiological and genetic analyses of Hepatitis C virus transmission among young/short- and long-term injecting drug users from Rio de Janeiro, Brazil. *J Clin Virol* **44**(3)**,** 200-6.

Pando, M. A., Bautista, C. T., Maulen, S., Duranti, R., Marone, R., Rey, J., Vignoles, M., Eirin, M. E., Biglione, M. M., Griemberg, G., Montano, S. M., Carr, J. K., Sanchez, J. L., and Avila, M. M. (2006). Epidemiology of human immunodeficiency virus, viral hepatitis (B and C), treponema pallidum, and human T-cell lymphotropic I/II virus among men who have sex with men in Buenos Aires, Argentina. *Sex Transm Dis* **33**(5)**,** 307-13.

Parfieniuk, A., Jaroszewicz, J., and Flisiak, R. (2007). Specifically targeted antiviral therapy for hepatitis C virus. *World J Gastroenterol* **13**(43)**,** 5673-81.

Parthiban, R., Shanmugam, S., Velu, V., Nandakumar, S., Dhevahi, E., Thangaraj, K., Nayak, H. K., Gupte, M. D., and Thyagarajan, S. P. (2009). Transmission of hepatitis C virus infection from asymptomatic mother to child in southern India. *Int J Infect Dis* **13**(6)**,** e394-400.

Pawlotsky, J. M. (1999). Hepatitis C virus (HCV) NS5A protein: role in HCV replication and resistance to interferon-alpha. *J Viral Hepat* **6**(Suppl 1)**,** 47-8.

Pawlotsky, J. M. (2000). Hepatitis C virus resistance to antiviral therapy. *Hepatology* **32**(5)**,** 889-96.

Pawlotsky, J. M. (2003a). Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antiviral Res* **59**(1)**,** 1-11.

Pawlotsky, J. M. (2003b). The nature of interferon-alpha resistance in hepatitis C virus infection. *Curr Opin Infect Dis* **16**(6)**,** 587-92.

Pawlotsky, J. M. (2003c). Hepatitis C virus genetic variability: pathogenic and clinical implications. *Clin Liver Dis* **7**(1)**,** 45-66.

Pawlotsky, J. M. (2004). Hepatitis C: it's a long way to new therapy, it's a long way to go. *Gastroenterology* **127**(5)**,** 1629-32.

Pawlotsky, J. M. (2005). Current and future concepts in hepatitis C therapy. *Semin Liver Dis* **25**(1)**,** 72-83.

Pawlotsky, J. M., Germanidis, G., Frainais, P. O., Bouvier, M., Soulier, A., Pellerin, M., and Dhumeaux, D. (1999). Evolution of the hepatitis C virus second envelope protein hypervariable region in chronically infected patients receiving alpha interferon therapy. *J Virol* **73**(8)**,** 6490-9.

Pawlotsky, J. M., Germanidis, G., Neumann, A. U., Pellerin, M., Frainais, P. O., and Dhumeaux, D. (1998). Interferon resistance of hepatitis C virus genotype 1b: relationship to nonstructural 5A gene quasispecies mutations. *J Virol* **72**(4)**,** 2795-805.

Pawlotsky, J. M., Tsakiris, L., Roudot-Thoraval, F., Pellet, C., Stuyver, L., Duval, J., and Dhumeaux, D. (1995). Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. *J Infect Dis* **171**(6)**,** 1607-10.

Payan, C., Roudot-Thoraval, F., Marcellin, P., Bled, N., Duverlie, G., Fouchard-Hubert, I., Trimoulet, P., Couzigou, P., Cointe, D., Chaput, C., Henquell, C., Abergel, A., Pawlotsky, J. M., Hezode, C., Coude, M., Blanchi, A., Alain, S., Loustaud-Ratti, V., Chevallier, P., Trepo, C., Gerolami, V., Portal, I., Halfon, P., Bourliere, M., Bogard, M., Plouvier, E., Laffont, C., Agius, G., Silvain, C., Brodard, V., Thiefin, G., Buffet-Janvresse, C., Riachi, G., Grattard, F., Bourlet, T., Stoll-Keller, F., Doffoel, M., Izopet, J., Barange, K., Martinot-Peignoux, M., Branger, M., Rosenberg, A., Sogni, P., Chaix, M. L., Pol, S., Thibault, V., Opolon, P., Charrois, A., Serfaty, L., Fouqueray, B., Grange, J. D., Lefrere, J. J., and Lunel-Fabiani, F. (2005). Changing of hepatitis C virus genotype patterns in France at the beginning of the third millenium: The GEMHEP GenoCII Study. *J Viral Hepat* **12**(4)**,** 405-13.

Pembrey, L., Newell, M. L., and Tovo, P. A. (2005). The management of HCV infected pregnant women and their children European paediatric HCV network. *J Hepatol* **43**(3)**,** 515-25.

Peng, J. S., Wang, X., Liu, M. Q., Zhou, D. J., Gong, J., Xu, H. M., Chen, J. P., Zhu, H. H., Zhou, W., and Ho, W. Z. (2008). Genetic variation of hepatitis C virus in a cohort of injection heroin users in Wuhan, China. *Virus Res* **135**(1)**,** 191-6.

Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D., and Pawlotsky, J. M. (2004). Structural biology of hepatitis C virus. *Hepatology* **39**(1)**,** 5-19.

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G., and Abrignani, S. (1998). Binding of hepatitis C virus to CD81. *Science* **282**(5390)**,** 938-41.

Ploss, A., Evans, M. J., Gaysinskaya, V. A., Panis, M., You, H., de Jong, Y. P., and Rice, C. M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* **457**(7231)**,** 882-6.

Pöhlmann, S., Zhang, J., Baribaud, F., Chen, Z., Leslie, G. J., Lin, G., Granelli-Piperno, A., Doms, R. W., Rice, C. M., and McKeating, J. A. (2003). Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J Virol* **77**(7)**,** 4070-80.

Ponamgi, S. P., Rahamathulla, S., Kumar, Y. N., Chandra, M., Lakshmi, N., Habibullah, C. M., and Khaja, M. N. (2009). Prevalence of hepatitis C virus (HCV) coinfection in HIV infected individuals in south India and characterization of HCV genotypes. *Indian J Med Microbiol* **27**(1)**,** 12-6.

Prince, A. M., Brotman, B., Huima, T., Pascual, D., Jaffery, M., and Inchauspe, G. (1992). Immunity in hepatitis C infection. *J Infect Dis* **165**(3)**,** 438-43.

Pukhrambam, P. D., Singh, N. B., Singh, L. H., and Devi Kh, S. (2007). Pattern of hepatitis C virus genotypes in HIV infected patients in RIMS, Hospital, Imphal, Manipur. *J Commun Dis* **39**(1)**,** 21-4.

Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C., and Harvey, P. H. (2001). The epidemic behavior of the hepatitis C virus. *Science* **292**(5525)**,** 2323-5.

Pybus, O. G., Cochrane, A., Holmes, E. C., and Simmonds, P. (2005). The hepatitis C virus epidemic among injecting drug users. *Infect Genet Evol* **5**(2)**,** 131-9.

Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H., and Rambaut, A. (2003). The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* **20**(3)**,** 381-7.

Ragni, M. V., and Belle, S. H. (2001). Impact of human immunodeficiency virus infection on progression to end-stage liver disease in individuals with hemophilia and hepatitis C virus infection. *J Infect Dis* **183**(7)**,** 1112-5.

Ramia, S., and Eid-Fares, J. (2006). Distribution of hepatitis C virus genotypes in the Middle East. *Int J Infect Dis* **23,** 23.

Rapicetta, M., Argentini, C., Spada, E., Dettori, S., Riccardi, M. P., and Toti, M. (2000). Molecular evolution of HCV genotype 2c persistent infection following mother-to-infant transmission. *Arch Virol* **145**(5)**,** 965-77.

Ray, S. C., Arthur, R. R., Carella, A., Bukh, J., and Thomas, D. L. (2000). Genetic epidemiology of hepatitis C virus throughout Egypt. *J Infect Dis* **182**(3)**,** 698-707.

Re, V., Contigiani, M., Yoshida, C. F., and Lampe, E. (2007). Identification of hepatitis C virus subtype 2c by sequencing analysis in patients from Cordoba, Argentina. *Mem Inst Oswaldo Cruz* **102**(8)**,** 995-8.

Re, V., Gallego, S., Farias, A., Barbas, G., Kremer, L., Diaz, M. P., and Contigiani, M. (2008). Hepatitis C and HIV coinfection in central region of Argentina: prevalence, genotype characterization and risk factors. *Enferm Infecc Microbiol Clin* **26**(7)**,** 423-5.

Reindollar, R. W. (1999). Hepatitis C and the correctional population. *Am J Med* **107**(6B)**,** 100S-103S.

Reiser, M., Hinrichsen, H., Benhamou, Y., Reesink, H. W., Wedemeyer, H., Avendano, C., Riba, N., Yong, C. L., Nehmiz, G., and Steinmann, G. G. (2005). Antiviral efficacy of NS3-serine protease inhibitor BILN-2061 in patients with chronic genotype 2 and 3 hepatitis C. *Hepatology* **41**(4)**,** 832-5.

Rhodes, A. G., Taxman, F. S., Friedmann, P. D., and Cropsey, K. L. (2008). HCV in incarcerated populations: an analysis of gender and criminality on risk. *J Psychoactive Drugs* **40**(4)**,** 493-501.

Rhodes, T., Platt, L., Judd, A., Mikhailova, L. A., Sarang, A., Wallis, N., Alpatova, T., Hickman, M., and Parry, J. V. (2005). Hepatitis C virus infection, HIV co-infection, and associated risk among injecting drug users in Togliatti, Russia. *Int J STD AIDS* **16**(11)**,** 749-54.

Richardson, D., Fisher, M., and Sabin, C. A. (2008). Sexual transmission of hepatitis C in MSM may not be confined to those with HIV infection. *J Infect Dis* **197**(8)**,** 1213-4, author reply 1214-5.

Rinaldi, A. (2006). The phantom menace. *EMBO Rep* **7**(1)**,** 14-7.

Robertson, B., Myers, G., Howard, C., Brettin, T., Bukh, J., Gaschen, B., Gojobori, T., Maertens, G., Mizokami, M., Nainan, O., Netesov, S., Nishioka, K., Shin i, T., Simmonds, P., Smith, D., Stuyver, L., and Weiner, A. (1998). Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. *Arch Virol* **143**(12)**,** 2493-503.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4)**,** 406-25.

Samimi-Rad, K., Nategh, R., Malekzadeh, R., Norder, H., and Magnius, L. (2004). Molecular epidemiology of hepatitis C virus in Iran as reflected by phylogenetic analysis of the NS5B region. *J Med Virol* **74**(2)**,** 246-52.

Sandres-Saune, K., Deny, P., Pasquier, C., Thibaut, V., Duverlie, G., and Izopet, J. (2003). Determining hepatitis C genotype by analyzing the sequence of the NS5b region. *J Virol Methods* **109**(2)**,** 187-93.

Savvas, S. P., Koskinas, J., Sinani, C., Hadziyannis, A., Spanou, F., and Hadziyannis, S. J. (2005). Changes in epidemiological patterns of HCV infection and their impact on liver disease over the last 20 years in Greece. *J Viral Hepat* **12**(5)**,** 551-7.

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., and Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *Embo J* **21**(19)**,** 5017-25.

Schroter, M., Zollner, B., Laufs, R., and Feucht, H. H. (2004). Changes in the prevalence of hepatitis C virus genotype among injection drug users: a highly dynamic process. *J Infect Dis* **190**(6)**,** 1199-200; author reply 1200-1.

Seeff, L. B. (2002). Natural history of chronic hepatitis C. *Hepatology* **36**(5 Suppl 1)**,** S35-46.

Sentandreu, V., Jimenez-Hernandez, N., Torres-Puente, M., Bracho, M. A., Valero, A., Gosalbes, M. J., Ortega, E., Moya, A., and Gonzalez-Candelas, F. (2008). Evidence of recombination in intrapatient populations of hepatitis C virus. *PLoS One* **3**(9)**,** e3239.

Shapatava, E., Nelson, K. E., Tsertsvadze, T., and del Rio, C. (2006). Risk behaviors and HIV, hepatitis B, and hepatitis C seroprevalence among injection drug users in Georgia. *Drug Alcohol Depend* **82 Suppl 1,** S35-8.

Shebl, F. M., El-Kamary, S. S., Saleh, D. A., Abdel-Hamid, M., Mikhail, N., Allam, A., El-Arabi, H., Elhenawy, I., El-Kafrawy, S., El-Daly, M., Selim, S., El-Wahab, A. A., Mostafa, M., Sharaf, S., Hashem, M., Heyward, S., Stine, O. C., Magder, L. S., Stoszek, S., and Strickland, G. T. (2009). Prospective cohort study of mother-to-infant infection and clearance of hepatitis C in rural Egyptian villages. *J Med Virol* **81**(6)**,** 1024-31.

Shepard, C. W., Finelli, L., and Alter, M. J. (2005). Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* **5**(9)**,** 558-67.

Sherman, K. E., Fleischer, R., Laessig, K., Murray, J., Tauber, W., and Birnkrant, D. (2007). Development of novel agents for the treatment of chronic hepatitis C infection: Summary of the FDA Antiviral Products Advisory Committee recommendations. *Hepatology* **46**(6)**,** 2014-20.

Shih, J. W., Mur, J. I., and Alter, H. J. (1986). Non-A, non-B hepatitis: advances and unfulfilled expectations of the first decade. *Prog Liver Dis* **8,** 433-52.

Simmonds, P. (1999). Viral heterogeneity of the hepatitis C virus. *J Hepatol* **31 Suppl 1,** 54-60.

Simmonds, P., Bukh, J., Combet, C., Deleage, G., Enomoto, N., Feinstone, S., Halfon, P., Inchauspe, G., Kuiken, C., Maertens, G., Mizokami, M., Murphy, D. G., Okamoto, H., Pawlotsky, J. M., Penin, F., Sablon, E., Shin, I. T., Stuyver, L. J., Thiel, H. J., Viazov, S., Weiner, A. J., and Widell, A. (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**(4)**,** 962-73.

Simmonds, P., Holmes, E. C., Cha, T. A., Chan, S. W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J., and Urdea, M. S. (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol* **74**(Pt 11)**,** 2391-9.

Simmonds, P., Smith, D. B., McOmish, F., Yap, P. L., Kolberg, J., Urdea, M. S., and Holmes, E. C. (1994). Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J Gen Virol* **75**(Pt 5)**,** 1053-61.

Smith, D. B., and Simmonds, P. (1997). Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome. *J Mol Evol* **45**(3)**,** 238-46.

Soriano, V., Peters, M. G., and Zeuzem, S. (2009). New therapies for hepatitis C virus infection. *Clin Infect Dis* **48**(3)**,** 313-20.

Spaulding, A. C., Weinbaum, C. M., Lau, D. T., Sterling, R., Seeff, L. B., Margolis, H. S., and Hoofnagle, J. H. (2006). A framework for management of hepatitis C in prisons. *Ann Intern Med* **144**(10)**,** 762-9.

Stuyver, L., van Arnhem, W., Wyseur, A., Hernandez, F., Delaporte, E., and Maertens, G. (1994). Classification of hepatitis C viruses based on phylogenetic analysis of the envelope 1 and nonstructural 5B regions and identification of five additional subtypes. *Proc Natl Acad Sci U S A* **91**(21)**,** 10134-8.

Stuyver, L., Wyseur, A., van Arnhem, W., Hernandez, F., and Maertens, G. (1996). Second-generation line probe assay for hepatitis C virus genotyping. *J Clin Microbiol* **34**(9)**,** 2259-66.

Sulkowski, M. S. (2008). Viral hepatitis and HIV coinfection. *J Hepatol* **48**(2)**,** 353-67.

Sutton, A. J., Hope, V. D., Mathei, C., Mravcik, V., Sebakova, H., Vallejo, F., Suligoi, B., Brugal, M. T., Ncube, F., Wiessing, L., and Kretzschmar, M. (2008). A comparison between the force of infection estimates for blood-borne viruses in injecting drug user populations across the European Union: a modelling study. *J Viral Hepat* **15**(11)**,** 809-16.

Sweeting, M. J., De Angelis, D., Brant, L. J., Harris, H. E., Mann, A. G., and Ramsay, M. E. (2007). The burden of hepatitis C in England. *J Viral Hepat* **14**(8)**,** 570-6.

Sweeting, M. J., Hope, V. D., Hickman, M., Parry, J. V., Ncube, F., Ramsay, M. E., and De Angelis, D. (2009). Hepatitis C infection among injecting drug users in England and Wales (1992-2006): there and back again? *Am J Epidemiol*.

Sy, T., and Jamal, M. M. (2006). Epidemiology of hepatitis C virus (HCV) infection. *Int J Med Sci* **3**(2)**,** 41-6. Epub 2006 Apr 1.

Tallo, T., Norder, H., Tefanova, V., Krispin, T., Schmidt, J., Ilmoja, M., Orgulas, K., Pruunsild, K., Priimagi, L., and Magnius, L. O. (2007). Genetic characterization of hepatitis C virus strains in Estonia: fluctuations in the predominating subtype with time. *J Med Virol* **79**(4)**,** 374-82.

Tamalet, C., Colson, P., Tissot-Dupont, H., Henry, M., Tourres, C., Tivoli, N., Botta, D., Ravaux, I., Poizot-Martin, I., and Yahi, N. (2003). Genomic and phylogenetic analysis of hepatitis C virus isolates: a survey of 535 strains circulating in southern France. *J Med Virol* **71**(3)**,** 391-8.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**(8)**,** 1596-9.

Tan, Y., Wei, Q. H., Chen, L. J., Chan, P. C., Lai, W. S., He, M. L., Kung, H. F., and Lee, S. S. (2008). Molecular epidemiology of HCV monoinfection and HIV/HCV coinfection in injection drug users in Liuzhou, Southern China. *PLoS ONE* **3**(10)**,** e3608.

Tanaka, Y., Agha, S., Saudy, N., Kurbanov, F., Orito, E., Kato, T., Abo-Zeid, M., Khalaf, M., Miyakawa, Y., and Mizokami, M. (2004). Exponential spread of hepatitis C virus genotype 4a in Egypt. *J Mol Evol* **58**(2)**,** 191-5.

Taylor, D. R. (2000). Hepatitis C virus: evasion of the interferon-induced antiviral response. *J Mol Med* **78**(4)**,** 182-90.

Thimme, R., Bukh, J., Spangenberg, H. C., Wieland, S., Pemberton, J., Steiger, C., Govindarajan, S., Purcell, R. H., and Chisari, F. V. (2002). Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. *Proc Natl Acad Sci U S A* **99**(24)**,** 15661-8.

Thompson, A. J., and McHutchison, J. G. (2009). Antiviral resistance and specifically targeted therapy for HCV (STAT-C). *J Viral Hepat* **16**(6)**,** 377-87.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24)**,** 4876-82.

Thomson, B. J. (2009). Hepatitis C virus: the growing challenge. *Br Med Bull* **89,** 153-67.

Timm, J., Neukamm, M., Kuntzen, T., Kim, A. Y., Chung, R. T., Brander, C., Lauer, G. M., Walker, B. D., and Allen, T. M. (2007). Characterization of full-length hepatitis C virus genotype 4 sequences. *J Viral Hepat* **14**(5)**,** 330-7.

Trepo, C., and Pradat, P. (1999). Hepatitis C virus infection in Western Europe. *J Hepatol* **31 Suppl 1,** 80-3.

Urbanus, A. T., van de Laar, T. J., Stolte, I. G., Schinkel, J., Heijman, T., Coutinho, R. A., and Prins, M. (2009). Hepatitis C virus infections among HIV-infected men who have sex with men: an expanding epidemic. *AIDS* **23**(12)**,** F1-7.

van Asten, L., Verhaest, I., Lamzira, S., Hernandez-Aguado, I., Zangerle, R., Boufassa, F., Rezza, G., Broers, B., Robertson, J. R., Brettle, R. P., McMenamin, J., Prins, M., Cochrane, A., Simmonds, P., Coutinho, R. A., and Bruisten, S. (2004). Spread of hepatitis C virus among European injection drug users infected with HIV: a phylogenetic analysis. *J Infect Dis* **189**(2)**,** 292-302.

van de Laar, T., Pybus, O., Bruisten, S., Brown, D., Nelson, M., Bhagani, S., Vogel, M., Baumgarten, A., Chaix, M. L., Fisher, M., Gotz, H., Matthews, G. V., Neifer, S., White, P., Rawlinson, W., Pol, S., Rockstroh, J., Coutinho, R., Dore, G. J., Dusheiko, G. M., and Danta, M. (2009). Evidence of a large, international network of HCV transmission in HIV-positive men who have sex with men. *Gastroenterology* **136**(5)**,** 1609-17.

van de Laar, T. J., Langendam, M. W., Bruisten, S. M., Welp, E. A., Verhaest, I., van Ameijden, E. J., Coutinho, R. A., and Prins, M. (2005). Changes in risk behavior and dynamics of hepatitis C virus infections among young drug users in Amsterdam, the Netherlands. *J Med Virol* **77**(4)**,** 509-18.

van de Laar, T. J., van der Bij, A. K., Prins, M., Bruisten, S. M., Brinkman, K., Ruys, T. A., van der Meer, J. T., de Vries, H. J., Mulder, J. W., van Agtmael, M., Jurriaans, S., Wolthers, K. C., and Coutinho, R. A. (2007). Increase in HCV incidence among men who have sex with men in Amsterdam most likely caused by sexual transmission. *J Infect Dis* **196**(2)**,** 230-8.

Vandelli, C., Renzo, F., Romano, L., Tisminetzky, S., De Palma, M., Stroffolini, T., Ventura, E., and Zanetti, A. (2004). Lack of evidence of sexual transmission of hepatitis C among monogamous couples: results of a 10-year prospective follow-up study. *Am J Gastroenterol* **99**(5)**,** 855-9.

Varaklioti, A., Vassilaki, N., Georgopoulou, U., and Mavromara, P. (2002). Alternate translation occurs within the core coding region of the hepatitis C viral genome. *J Biol Chem* **277**(20)**,** 17713-21. Epub 2002 Mar 7.

Verbeeck, J., Maes, P., Lemey, P., Pybus, O. G., Wollants, E., Song, E., Nevens, F., Fevery, J., Delport, W., Van der Merwe, S., and Van Ranst, M. (2006). Investigating the origin and spread of hepatitis C virus genotype 5a. *J Virol* **80**(9)**,** 4220-6.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H. G., Mizokami, M., Bartenschlager, R., and Liang, T. J. (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**(7)**,** 791-6.

Walewski, J. L., Keller, T. R., Stump, D. D., and Branch, A. D. (2001). Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**(5)**,** 710-21.

Wang, Y., Xia, X., Li, C., Maneekarn, N., Xia, W., Zhao, W., Feng, Y., Kung, H. F., Fu, Y., and Lu, L. (2009). A new HCV genotype 6 subtype designated 6v was confirmed with three complete genome sequences. *J Clin Virol* **44**(3)**,** 195-9.

Wasley, A., and Alter, M. J. (2000). Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin Liver Dis* **20**(1)**,** 1-16.

Watson, J. P., Al-Mardini, H., Awadh, S., Ukabam, S., and Record, C. O. (1999). Hepatitis C virus genotypes in a cohort of Middle Eastern patients. *Ann Saudi Med* **19**(5)**,** 410-2.

Webster, D. P., Klenerman, P., Collier, J., and Jeffery, K. J. (2009). Development of novel treatments for hepatitis C. *Lancet Infect Dis* **9**(2)**,** 108-17.

Weinbaum, C. M., Sabin, K. M., and Santibanez, S. S. (2005). Hepatitis B, hepatitis C, and HIV in correctional populations: a review of epidemiology and prevention. *AIDS* **19 Suppl 3,** S41-6.

Weiner, A. J., Christopherson, C., Hall, J. E., Bonino, F., Saracco, G., Brunetto, M. R., Crawford, K., Marion, C. D., Crawford, K. A., Venkatakrishna, S., and et al. (1991). Sequence variation in hepatitis C viral isolates. *J Hepatol* **13 Suppl 4,** S6-14.

Weiner, A. J., Paliard, X., Selby, M. J., Medina-Selby, A., Coit, D., Nguyen, S., Kansopon, J., Arian, C. L., Ng, P., Tucker, J., Lee, C. T., Polakos, N. K., Han, J., Wong, S., Lu, H. H., Rosenberg, S., Brasky, K. M., Chien, D., Kuo, G., and Houghton, M. (2001). Intrahepatic genetic inoculation of hepatitis C virus RNA confers cross-protective immunity. *J Virol* **75**(15)**,** 7142-8.

Weiner, A. J., Thaler, M. M., Crawford, K., Ching, K., Kansopon, J., Chien, D. Y., Hall, J. E., Hu, F., and Houghton, M. (1993). A unique, predominant hepatitis C virus variant found in an infant born to a mother with multiple variants. *J Virol* **67**(7)**,** 4365-8.

Xia, X., Lu, L., Tee, K. K., Zhao, W., Wu, J., Yu, J., Li, X., Lin, Y., Mukhtar, M. M., Hagedorn, C. H., and Takebe, Y. (2008b). The unique HCV genotype distribution and the discovery of a novel subtype 6u among IDUs co-infected with HIV-1 in Yunnan, China. *J Med Virol* **80**(7)**,** 1142-52.

Xia, X., Zhao, W., Tee, K. K., Feng, Y., Takebe, Y., Li, Q., Pybus, O. G., and Lu, L. (2008a). Complete genome sequencing and phylogenetic analysis of HCV isolates from China reveals a new subtype, designated 6u. *J Med Virol* **80**(10)**,** 1740-6.

Xu, Z., Choi, J., Yen, T. S., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M. J., and Ou, J. (2001). Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *Embo J* **20**(14)**,** 3840-8.

Yagnik, A. T., Lahm, A., Meola, A., Roccasecca, R. M., Ercole, B. B., Nicosia, A., and Tramontano, A. (2000). A model for the hepatitis C virus envelope glycoprotein E2. *Proteins* **40**(3)**,** 355-66.

Zalumas, J. C., and Rose, C. D. (2003). Hepatitis C and HIV in incarcerated populations: fights, bites, searches, and syringes! *J Assoc Nurses AIDS Care* **14**(5 Suppl)**,** 108S-115S.

Zein, N. N. (2000). Clinical significance of hepatitis C virus genotypes. *Clin Microbiol Rev* **13**(2)**,** 223-35.

Zein, N. N., Rakela, J., Krawitt, E. L., Reddy, K. R., Tominaga, T., and Persing, D. H. (1996). Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. *Ann Intern Med* **125**(8)**,** 634-9.

Zeisel, M. B., Koutsoudakis, G., Schnober, E. K., Haberstroh, A., Blum, H. E., Cosset, F. L., Wakita, T., Jaeck, D., Doffoel, M., Royer, C., Soulier, E., Schvoerer, E., Schuster, C., Stoll-Keller, F., Bartenschlager, R., Pietschmann, T., Barth, H., and Baumert, T. F. (2007). Scavenger receptor class B type I is a key host factor for hepatitis C virus

infection required for an entry step closely linked to CD81. *Hepatology* **46**(6)**,** 1722-31.

Zeuzem, S. (2004). Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well? *Ann Intern Med* **140**(5)**,** 370-81.

Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D. R., Wieland, S. F., Uprichard, S. L., Wakita, T., and Chisari, F. V. (2005). Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A* **102**(26)**,** 9294-9.

Zhou, D. X., Tang, J. W., Chu, I. M., Cheung, J. L., Tang, N. L., Tam, J. S., and Chan, P. K. (2006). Hepatitis C virus genotype distribution among intravenous drug user and the general population in Hong Kong. *J Med Virol* **78**(5)**,** 574-81.

Appendices

I. List of abbreviations

AIDS, acquired immune deficiency syndrome
HBV, hepatitis B virus
HCC, hepatocellular carcinoma
HCV, hepatitis C virus
HIV, human immunodeficiency virus
ER, endoplasmic reticulum
IRES, internal ribosome entry site
ISDR, interferon sensitivity-determining region
LDL, low-density lipoproteins
MSM, men who have sex with men
NANBH, non-A, non-B hepatitis
NCR, non-coding region
NJ, Neighbour-joining
NS, non-structural
ORF, open reading frame
RdRp, RNA-dependent RNA polymerase
SVR, sustained virological response
VLDL, very-low-density lipoproteins

II. Published work

A. Publications in peer-reviewed scientific journals [1]

1. Hadjinicolaou, A. V.; Farcas, G. A.; Demetriou, V. L.; Mazzulli, T.; Poutanen, S. M.; Willey, B.; Low, D. E.; Butany, J.; Asa, S. L.; Kain, K. C.; Kostrikis, L. G. Development of a molecular-beacon-based multi-allelic real-time RT-PCR assay for the detection of human coronavirus causing severe acute respiratory syndrome (SARS-CoV): A general methodology for detecting rapidly mutating viruses. *Arch Virol* In press.

2. Demetriou, V. L.; van de Vijver, D. A.; Kousiappa, I.; Balotta, C.; Bonaventura, C.; Grossman, Z.; Jørgensen, L. B.; Lepej, S. Z.; Levy, I.; Nielsen, C.; Paraskevis, D.; Poljak, M.; Roman, F.; Ruiz, L.; Schmidt, J. C.; Vandamme, A. M.; Van Laethem, K.; Vercauteren, J.; Kostrikis, L. G. Cellular HIV-1 DNA levels in drug sensitive strains are equivalent to those in drug resistant strains in newly-diagnosed patients in Europe. *PLoS ONE* **2010**; 5(6):e10976.

3. Demetriou, V. L.; van de Vijver, D. A.; Hezka, J.; The Cyprus IVDU Network; Kostrikis, L. G. Hepatitis C infection among intravenous drug users attending therapy programmes in Cyprus. *J Med Virol* **2010**; 82(2):263-70.

4. Hadjinicolaou, A. V.; Demetriou, V. L.[2]; Hezka, J.; Beyer, W.; Hadfield, T. L.; Kostrikis, L. G. Use of molecular beacons and multi-allelic real-time PCR for detection of and discrimination between virulent *Bacillus anthracis* and other *Bacillus* isolates. *J Microbiol Methods* **2009**; 78(1):45-53.

5. Hadjinicolaou, A. V.; Demetriou, V. L.[2]; Emmanuel, M. A.; Kakoyiannis, C. K.; Kostrikis, L. G. Molecular beacon-based real-time PCR detection of primary isolates of *Salmonella* Typhimurium and *Salmonella* Enteritidis in environmental and clinical samples. *BMC Microbiol* **2009**; 9:97-110.

[1] Copies of publications relevant to the thesis are attached after the Appendices.
[2] Co-first author contribution.

6. Demetriou, V. L.; van de Vijver, D. A.; The Cyprus HCV Network; Kostrikis, L. G. Molecular epidemiology of hepatitis C infection in Cyprus: evidence of polyphyletic infection. *J Med Virol* **2009**; 81(2):238-48.

7. Demetriou, V. L.; Kyriakou, E.; Kostrikis, L. G. Near-full genome characterisation of two natural intergenotypic 2k/1b recombinant Hepatitis C virus isolates. *Submitted*.

8. Demetriou, V. L.; the Cyprus HCV Network; Kostrikis, L. G. Molecular epidemiology of hepatitis C infection in Cyprus within the general population and high-risk cohorts: Near-full genome characterisation of unclassified strains within genotypes 1 and 4. *Submitted*.

B. Abstracts in international scientific conferences

1. Hadjinicolaou A. V., Demetriou V. L., Kostrikis L.G. Molecular Beacon Probes: A New Tool for Fast and Accurate Detection of Infectious Agents. *17th International Student Congress of Medical Sciences,* 8-11 June **2010**, Groningen, the Netherlands.[3]

2. Demetriou V. L., Kousiappa I., Kostrikis L. G. Molecular Epidemiology of HIV-1 and HCV in Co-Infected Individuals in Cyprus. *6th International Workshop on HIV & Hepatitis Co-Infection,* 31 May - 2 June **2010**, Tel Aviv, Israel.

3. Demetriou V. L., Kostrikis L. G. Implications of HCV Natural Genetic Diversity on HCV NS5B Inhibitor NM283. *16th International Symposium on HIV and Emerging Infectious Diseases*, 24-26 March **2010**, Marseille, France.

4. Demetriou V. L., van de Vijver D. A., Kostrikis L. G. Molecular Characterisation of HCV strains from Cypriot IV Drug Users Reveals Multiple Introductions and

[3] Awarded first prize in the Medical Microbiology section.

Epidemiologically Related Transmission Clusters. *8th European HIV Drug Resistance Workshop*, 17-19 March **2010**, Sorrento, Italy.

5. Demetriou V. L., Loizidou E. Z.**,** Kostrikis L. G. Implications of HCV Natural Genetic Diversity on HCV RNA Polymerase Nucleoside Inhibitor NM283. *8th European HIV Drug Resistance Workshop*, 17-19 March **2010**, Sorrento, Italy.

6. Demetriou V. L., van de Vijver D. A. M. C., Hezka J., Georgiades N., Savvopoulou N., Charilaou C. C., Argyriou A., Pavlou T., Platritis K., Veresies K. L.**,** Kostrikis L. G. Molecular Epidemiology of Hepatitis C Infection among Intravenous Drug Users Seeking Therapy in Cyprus. *16th International Symposium on Hepatitis C Virus and Related Viruses*, 3-7 October **2009**, Nice, France.

7. Demetriou V. L., Vounou-Hadjigeorgiou E., Evgeniou A., Koliou-Mazeri M., Papakyriakou P., Petsas L., Potamitis G., Cheimonidis S., Kostrikis L. G. Genetic Analysis of the Hepatitis C Virus (HCV) in Patients in Cyprus: Discovery of Genotypic Heterogeneity and Two Recombinant Strains. *27th Panhellenic Gastroenterology Conference*, 4-7 October **2007**, Thessaloniki, Greece.

8. Demetriou V. L., The Cyprus Hepatitis C Network, Kostrikis L. G. Epidemiological Analysis of the Hepatitis C Infection in Cyprus. *12th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology*, 11-15 September **2006**, Athens, Greece.

Journal of Medical Virology 82:263–270 (2010)

Hepatitis C Infection Among Intravenous Drug Users Attending Therapy Programs in Cyprus

Victoria L. Demetriou,[1] David A.M.C. van de Vijver,[2] Johana Hezka,[1] the Cyprus IVDU Network, and Leondios G. Kostrikis[1]*

[1]*Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus*
[2]*Department of Virology, Erasmus MC, University Medical Centre, Rotterdam, the Netherlands*

The most high-risk population for HCV transmission worldwide today are intravenous drug users. HCV genotypes in the general population in Cyprus demonstrate a polyphyletic infection and include subtypes associated with intravenous drug users. The prevalence of HCV, HBV, and HIV infection, HCV genotypes and risk factors among intravenous drug users in Cyprus were investigated here for the first time. Blood samples and interviews were obtained from 40 consenting users in treatment centers, and were tested for HCV, HBV, and HIV antibodies. On the HCV-positive samples, viral RNA extraction, RT-PCR and sequencing were performed. Phylogenetic analysis determined subtype and any relationships with database sequences and statistical analysis determined any correlation of risk factors with HCV infection. The prevalence of HCV infection was 50%, but no HBV or HIV infections were found. Of the PCR-positive samples, eight (57%) were genotype 3a, and six (43%) were 1b. No other subtypes, recombinant strains or mixed infections were observed. The phylogenetic analysis of the injecting drug users' strains against database sequences observed no clustering, which does not allow determination of transmission route, possibly due to a limitation of sequences in the database. However, three clusters were discovered among the drug users' sequences, revealing small groups who possibly share injecting equipment. Statistical analysis showed the risk factor associated with HCV infection is drug use duration. Overall, the polyphyletic nature of HCV infection in Cyprus is confirmed, but the transmission route remains unknown. These findings highlight the need for harm-reduction strategies to reduce HCV transmission. ***J. Med. Virol. 82:263–270, 2010.*** © 2009 Wiley-Liss, Inc.

KEY WORDS: molecular epidemiology; HCV genotypes; NS5B region; Core-E1 region; phylogenetic analysis

INTRODUCTION

Hepatitis C is a life-shortening disease recognized as a major public health problem worldwide, infecting more than 170 million people globally [Sy and Jamal, 2006]. It is caused by a small, enveloped single-stranded, positive-sense RNA virus of the *Flaviviridae* family. The HCV genome reveals high genetic heterogeneity, leading to a proposed consensus of six genotypes and numerous closely related subtypes [Simmonds et al., 2005]. Genotype distribution differs by geographic region but also by year and mode of transmission, but the face of HCV epidemiology is being changed radically by globalization [Esteban et al., 2008].

The most important route of HCV transmission is through exposure to infected blood and until diagnostic blood screening was introduced in the early 1990s the virus was transmitted mainly through blood, blood products, hemodialysis and organ transplantations. In the western world currently, HCV infection occurs mainly through parenteral exposure, the most common mode of transmission being intravenous drug use through sharing of needles or other injecting equipment [Shepard et al., 2005; Sy and Jamal, 2006; Esteban et al., 2008]. Long-term users display an extremely high prevalence of HCV infection at over 60%, whereas users who have been practicing intravenous drug abuse

Collaborators—Members of the Cyprus IVDU network: Natasa Savvopoulou, Cyprus National Monitoring Centre for Drugs, Nicosia; Neoklis Georgiades, Cyprus National Monitoring Centre for Drugs, Nicosia; Charis C. Charilaou, Charilaou Biomedical Laboratory, Nicosia; Argyris Argyriou, Gephyra Substitution Treatment Programme, Nicosia and Anosis Detoxification Unit, Limassol; Tina Pavlou, Aghia Skepi Therapeutic Community, Nicosia; Kyriakos Platritis, Ploeghos Psychosocial Treatment Programme, Limassol; Kyriakos L. Veresies, Veresie Clinic, Larnaca.

*Correspondence to: Leondios G. Kostrikis, Laboratory of Biotechnology and Molecular Virology, Department of Biological Sciences, University of Cyprus, 75 Kallipoleos Avenue, PO Box 20537, 1678 Nicosia, Cyprus. E-mail: lkostrik@ucy.ac.cy

Accepted 21 September 2009

DOI 10.1002/jmv.21690

Published online in Wiley InterScience (www.interscience.wiley.com)

exhibit prevalence rates of 20–46%, which is still a significantly high rate [Shepard et al., 2005; Sweeting et al., 2009]. In general, large heterogeneity has been observed in the prevalence of HCV across different populations of injecting drug users [Hagan et al., 2007; Sutton et al., 2008].

In developed countries, the HCV subtypes that predominate among intravenous drug users are 1a and 3a, both showing exponential growth during the 20th century [Pybus et al., 2005; Esteban et al., 2008]. Subtype 3a has been associated significantly with transmission through intravenous drug use in industrialized countries. It is prevalent mainly in North and South America, Europe, and Australia, where intravenous drug use is common, and seems to have a common origin among injecting drug users [Pybus et al., 2005; Morice et al., 2006]. Lately genotype 4 strains, which have been restricted to north Africa, are also becoming increasingly prevalent among intravenous drug users, especially in southern Europe, and its introduction into this population seems to be more recent than that of 1a and 3a [van Asten et al., 2004; Esteban et al., 2008]. There is also one recombinant, the 2k/1b strain, which is associated with intravenous drug use, and has been found in Russia, Estonia, Uzbekistan, Ireland, and Cyprus [Kalinina et al., 2002; Moreau et al., 2006; Tallo et al., 2007; Kurbanov et al., 2008; Demetriou et al., 2009].

The molecular epidemiology of HCV infection in the general population of Cyprus was studied recently, revealing polyphyletic infection and a high genetic heterogeneity among strains on the island [Demetriou et al., 2009]. The presence of strains belonging to five of the six genotypes and putative isolates of the 2k/1b recombinant suggest multiple routes of introduction into the island from various geographical regions and by various routes of transmission. Cyprus is a small country, but unique with respect to its geographical position, as it lies between Europe, Africa, and Asia. It is very close to North Africa, where the prevalence of HCV is the highest in the world and Cyprus has a high rate of influx of immigrants from Eastern Europe and countries of the former Soviet Union. Also, as it is now a member state of the European Union, entry into Cyprus is easy, thus facilitating the introduction of new infectious diseases.

The molecular epidemiology of HCV, and also HBV and HIV infection specifically among a population of intravenous drug users in Cyprus was investigated in this study, following the recent discovery of subtypes associated with injecting drug use (1a, 3a, and 2k/1b) within the general population [Demetriou et al., 2009]. The aim of the study was to determine the prevalence of HCV, HBV, and HIV infection within this high-risk population, to determine the genotype distribution and investigate any phylogenetic relationships of strains found between themselves and with the strains found previously in the general population of Cyprus but also with the strains available in the Los Alamos HCV database (http://hcv.lanl.gov). The molecular epidemiology of hepatitis C infection among intravenous drug users in Cyprus has never been studied before and is presented for the first time, revealing genetic heterogeneity and multiple points of introduction of the virus within this high-risk population.

MATERIALS AND METHODS

Study Subjects and Sample Collection

The study subjects were consenting intravenous drug users seeking therapy in detoxification centers around the island at the time of sample collection. Two rounds of sampling were carried out in July and November 2008, covering detoxification centers treating intravenous drug users. The centers where the sampling took place were as follows: Aghia Skepi, a long-term inpatient therapeutic community which aims for withdrawal from substance abuse and social rehabilitation; Ghephyra, an outpatient substitution treatment program, for users who have failed with other therapeutic programs and buprenorphine is administered in parallel to psychosocial intervention and psychological and medical aid; Anosis, a long-term inpatient detoxification program, for abstinence from drug use and safe management of withdrawal symptoms through administration of analgesics and psychosocial support; Ploeghos, an outpatient psychosocial treatment program, for users who have failed with other therapeutic programs and buprenorphine is administered in parallel to psychosocial intervention and psychological and medical aid; Veresie Clinic, a clinic offering detoxification services through psychological and medical services, and perform naltrexone implant procedures.

An informed consent form was signed by each subject, a questionnaire was filled in with an interviewer, and blood samples were taken by qualified personnel as described previously [Demetriou et al., 2009]. All samples and questionnaires were coded at random with a laboratory identifier number so as not to reveal the personal identity of the study subjects. Forty individuals in total were included in this study, representing 87% of intravenous drug users requesting detoxification support at the time of sampling.

HCV, HBV, and HIV Tests

Immunoassays were performed on all samples for detection of HCV, HBV, HIV-1 and HIV-2 antibodies, using the AxSYM anti-HCV, HIV and HBsAg systems (Abbott Diagnostics, Chicago, IL).

RNA Extraction, RT-PCR and Sequencing in the Core-E1 and NS5B Regions

On all samples tested positive for HCV antibodies, RNA extraction, reverse transcription-PCR and sequencing were carried out in the regions Core-E1 and NS5B as described previously [Demetriou et al., 2009].

Subtyping and Phylogenetic Analysis

Subtyping of the strains found was done using the Oxford HCV Subtyping Tool v1.0 [de Oliveira et al., 2005]. Phylogenetic analysis was performed as described previously [Demetriou et al., 2009] to confirm the subtype of the strains and to investigate any possible phylogenetic relationship between the strains identified in this study from the intravenous drug users cohort and those identified in the general population of Cyprus [Demetriou et al., 2009]. Briefly, the sequences of both the Core-E1 region (14 sequences, 417 bp, positions 867–1283) and NS5B region (12 sequences, 405 bp, positions 8277–8681) were aligned and compared to reference strains from the Los Alamos database (http://hcv.lanl.gov) using the neighbor-joining method [Saitou and Nei, 1987] and the Kimura two-parameter distance estimation approach [Kimura, 1980] in MEGA v4 [Tamura et al., 2007]. The reliability of the phylogenetic clustering was evaluated using bootstrap analysis with 1,000 replicates [Felsenstein, 1985]. Bootstrap values above 70 were considered sufficient for subtype assignment. Clusters were determined by an arbitrary threshold of genetic distance (<0.05) [Lewis et al., 2008].

Additionally, for the purpose of a deeper investigation into possible phylogenetic relationships between the dataset obtained in this study and other HCV strains, each sequence was uploaded individually into the HCVBLAST tool on the Los Alamos HCV database website to discover the 50 most closely related sequences available in the database. Phylogenies for these sequences were constructed with MrBayes [Huelsenbeck and Ronquist, 2001] using a general time-reversible (GTR) model of nucleotide substitution with a proportion of invariant sites (ι) and gamma distribution of rates (Γ). The Monte–Carlo Markov Chain search was run for 5×10^6 generations, with trees sampled every 100th generation (with a burn-in of 50%) and a posterior consensus tree generated (from 25,000 trees). From this consensus tree, the posterior probability of nodes was used as phylogenetic support for each transmission cluster group. Cluster group size was determined using nodes with a posterior probability of 1.

Reference Sequences

The GenBank accession numbers for the reference sequences used in the phylogenetic analyses of the Core-E1 and NS5B regions are: AB031663, AF064490, AF169004, AF207752, AF238486, AF290978, AJ000009, AY051292, AY587845, D10988, D14853, D17763, D28917, D49374, D50409, D63821, D85516, D90208, DQ418786, DQ418787, DQ418789, EF108306, EF589160, EF589161, M62321, M67463, NC_004102, NC_009823, NC_009825, NC_009826, NC_009827, X76918, Y12083. The GenBank accession numbers for the sequences of the Cypriot strains of the general population used in the phylogenetic analysis for Core-E1 are EU684661–EU684737 and for NS5B are EU684591–EU684660.

Statistical Analysis

Variables were compared between hepatitis C positive and negative individuals using Mann–Whitney test for continuous variables and Chi-square test or Fisher-exact test where appropriate.

Nucleotide Sequence Accession Numbers

GenBank accession numbers for the sequences obtained in this study are GQ332540–GQ332553 for the Core-E1 sequences and GQ332554–GQ332565 for the NS5B sequences.

RESULTS

Demographic and Epidemiological Characteristics of the Study Subjects

The characteristics of the study subjects are listed in Table I, showing a comparison of those tested positive and those tested negative for HCV antibodies. Forty individuals in total were included in the study, of which the majority were male (85%) with a median age of 27 years (interquartile range 25–31) and of Cypriot nationality (64%). Other nationalities were either Greek or Eastern European. There was no statistical significance between the distribution of age or nationality between the HCV positive and negative individuals ($P = 0.61$ and 0.30 respectively). The results of the antibody testing showed that none of the intravenous drug users who took part in the study were infected with HIV or HBV at the time of sampling. However, 20 individuals of the 40 tested were positive for HCV antibodies, giving a prevalence of 50%.

The main risk factor which was statistically significant for HCV infection was the duration of injection drug use which, with a median duration of 10 years (IQR 5–14) was significantly higher in HCV-positive individuals as compared to a value of 6 years (IQR 1.5–8.5) in the HCV-negative group ($P = 0.02$). The association of the duration of intravenous drug use with the prevalence of HCV infection is seen in Figure 1. Another risk factor significantly different between the HCV positive and negative groups was a prison sentence. HCV-positive users reported a prison sentence more frequently than those who were negative ($P = 0.01$). All subjects reported to have practiced unprotected sex. Further analysis of known risk factors for HCV infection revealed no significant differences between the HCV positive and negative groups.

RT-PCR Sensitivity

Of the 20 samples found to be HCV positive from the AxSYM system, only 14 were PCR positive in the Core-E1 region and 12 were PCR positive in the NS5B region. PCR success rate was determined by comparing PCR results (positive or negative) with the results from the AxSYM anti-HCV system, which gives a value for the sample/cutoff rate ratio (S/CO) for each sample, where a value above 1 is considered positive. The S/CO values for the positive samples in this study were 4–124.

TABLE I. Characteristics of the Study Subjects

Characteristics[a]	Subjects (N = 40)	HCV +ve	HCV −ve
Age (years) (%)			
Median (IQR)	27 (25–31)	27 (24–34)	27 (25–29)
Gender (%)			
Male	35 (85)	16	19
Female	6 (15)	3	3
Nationality (%)			
Cypriot	25 (64)	10	15
Greek	8 (21)	4	4
Russian	2 (5)	2	0
Bulgarian	1 (3)	1	0
Georgian	1 (3)	1	0
Romanian	1 (3)	0	1
Not stated	1 (3)	0	1
Risk factors (%)			
Age at first injection (years) (IQR)	20 (17–23)	19.5 (16–21)	21 (18–25)
Duration of injecting drug use (years) (IQR)		10 (5–14)	6 (2–9)
Served a prison sentence	11 (28)	9	2
Shared injecting equipment	27 (68)	13	14
Can find sterile equipment easily	26 (65)	12	14
History of blood transfusion	3 (8)	3	0
History of surgical procedure	24 (60)	12	12
Tattoos	25 (63)	12	13
Have used syringes/needles abroad	10 (25)	7	3
Sexual practices (%)			
Heterosexual	35 (92)	17	18
Homosexual	1 (3)	1	0
Bisexual	2 (5)	0	2
Unprotected sexual practices	40 (100)	19	21
RT-PCR[b]	*Core-E1*	*NS5B*	
Positive	14	12	
Negative	6	8	
AxSYM anti-HCV S/CO values (IQR)[c]	*Core-E1*	*NS5B*	
PCR positive	86.8 (47.8–95.7)	88.1 (77.8–95.9)	
PCR negative	8.8 (4.6–25.7)	14.6 (5.5–45.6)	
Genotype[d]	*Core-E1*	*NS5B*	
1b	6	4	
3a	8	8	

[a]IQR, interquartile range.
[b]Only performed on samples which were HCV seropositive.
[c]AxSYM anti-HCV system, the HCV antibody test used in this study; S/CO, signal/cutoff rate ratio is the result obtained from the HCV antibody test.
[d]Only obtained for samples which were PCR positive.

A Mann–Whitney test showed that the PCR positive samples were those with significantly higher S/CO values than the PCR negative samples for both regions tested, as seen in Table I ($P = 0.001$).

Core-E1 and NS5B Phylogenetic Analyses

Phylogenetic trees of the Core-E1 and NS5B sequences obtained in this study with reference sequences and the sequences obtained from the general population infected with HCV in Cyprus are seen in Figures 2 and 3, respectively. The results of the subtyping showed that the strains of this study group fell into only two subtypes. Eight samples (57%) belong to subtype 3a, and six samples (43%) belong to subtype 1b, with bootstrap values higher than 98. Two of the 3a strains are not seen in the NS5B tree, as they were PCR negative in this region. There was concurrence for all other samples between the two regions.

No clusters were observed between the sequences from the intravenous dug users and those of the general population, however three small clusters were seen within the injecting drug users group with small genetic distances (<0.05) and high bootstrap values. These clusters are between samples CYIDU06, CYIDU07, and CYIDU33 within genotype 1b, CYIDU24 and CYIDU37 also within 1b, and CYIDU20, CYIDU26, and CYIDU27 within genotype 3a. The two subjects clustering together in the 1b subtype (CYIDU24 and CYIDU37) interestingly reported the use of speedball. Only one other individual reported this, CYIDU02, but they tested negative for HCV antibodies.

The Bayesian analysis of the Cypriot intravenous drug users sequences with the most genetically related

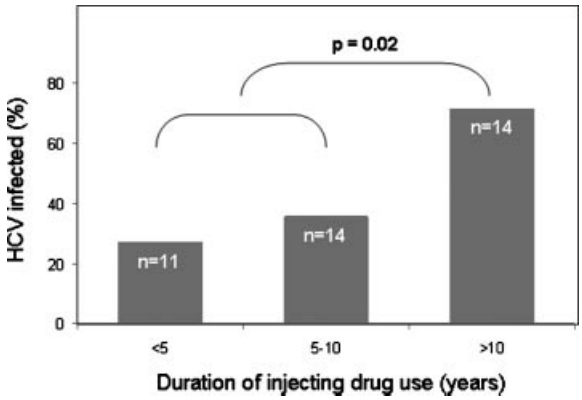

Fig. 1. Bar chart indicating the percentage prevalence of HCV seropositive users in three groups of different intravenous drug use duration. In the group of individuals who had been users for up to 5 years, 3 out of 11 were anti-HCV positive (prevalence is 27%); within the individuals who had practiced IDU for 5–10 years, 5 out of 14 were anti-HCV positive (prevalence is 36%); within the individuals who had been practicing IDU for over 10 years, 10 out of 14 were positive (prevalence is 71%). Intravenous drug users who had been practicing for more than 10 years were significantly at higher risk of being HCV seropositive than short-term users ($P = 0.02$).

database sequences, which resulted from the BLAST search, identified no significant clustering, as indicated by posterior probabilities that did not exceed 0.90 (data not shown).

DISCUSSION

In this study, the prevalence and molecular epidemiology among intravenous drug users seeking therapy in Cyprus is presented. Forty consenting individuals took part in this work, providing demographic and epidemiological characteristics as well as information on risk groups and drug use behavior. The study subjects were predominantly young men, and mostly Cypriots. The nationalities of the subjects who were not Cypriot demonstrate the influx of young people from Eastern Europe to the island and its strong association with Greece.

The results of the study showed that within the group tested, the prevalence of HCV infection was 50%, which is comparable to or lower than known prevalence rates in intravenous drug users [Shepard et al., 2005; van de

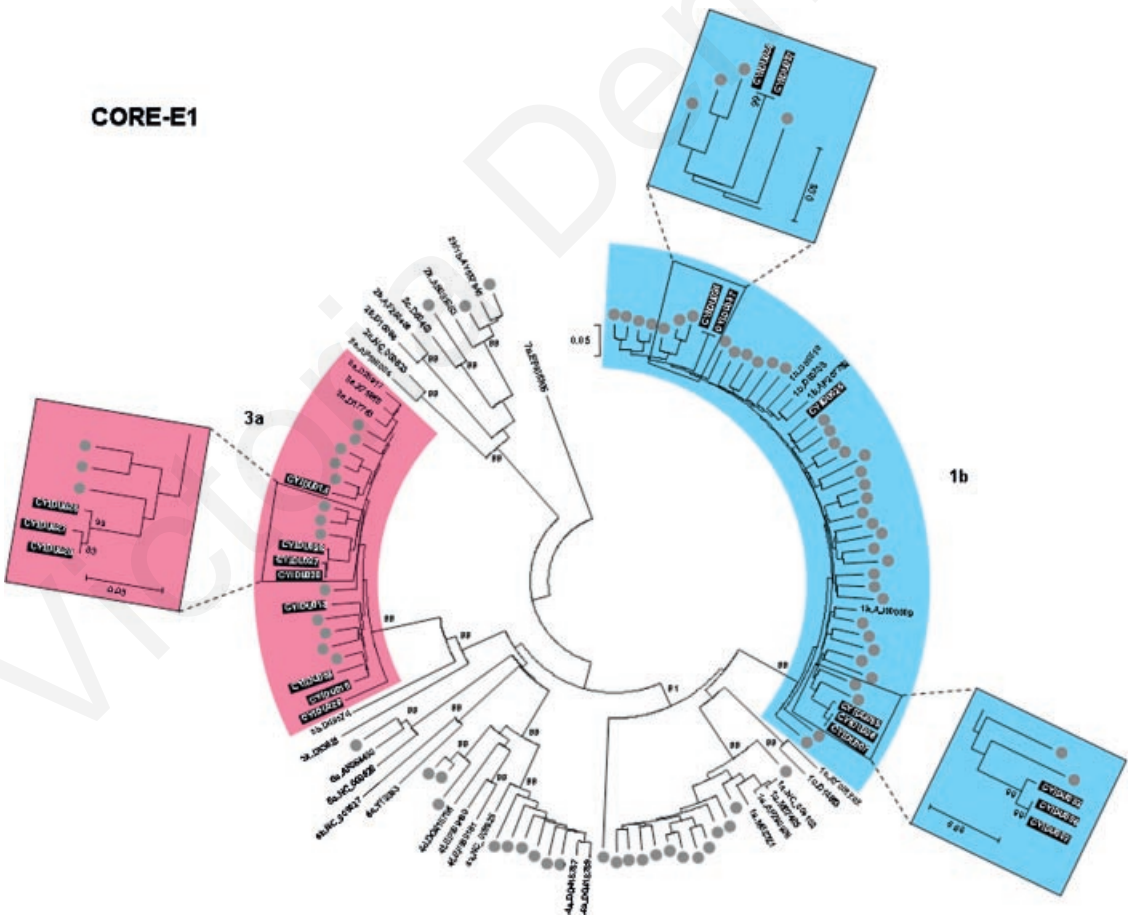

Fig. 2. Neighbor-joining tree of Core-E1 sequences, constructed as described in the "Materials and Methods" section. The intravenous drug users' strains are indicated with their laboratory codes and highlighted in black with white font. The reference strains are written in black font. The strains of the general population are indicated with gray circles. The subtypes in which the injecting drug users' strains fall are colored: 1b is blue and 3a is pink. The observed clusters of intravenous drug users' strains are indicated with a square and magnified for better viewing. The numbers indicated at the nodes are percentage bootstrap support values for 1,000 replicates. The divergence between any two sequences is obtained by summing the branch length using the scale at the top of the tree.

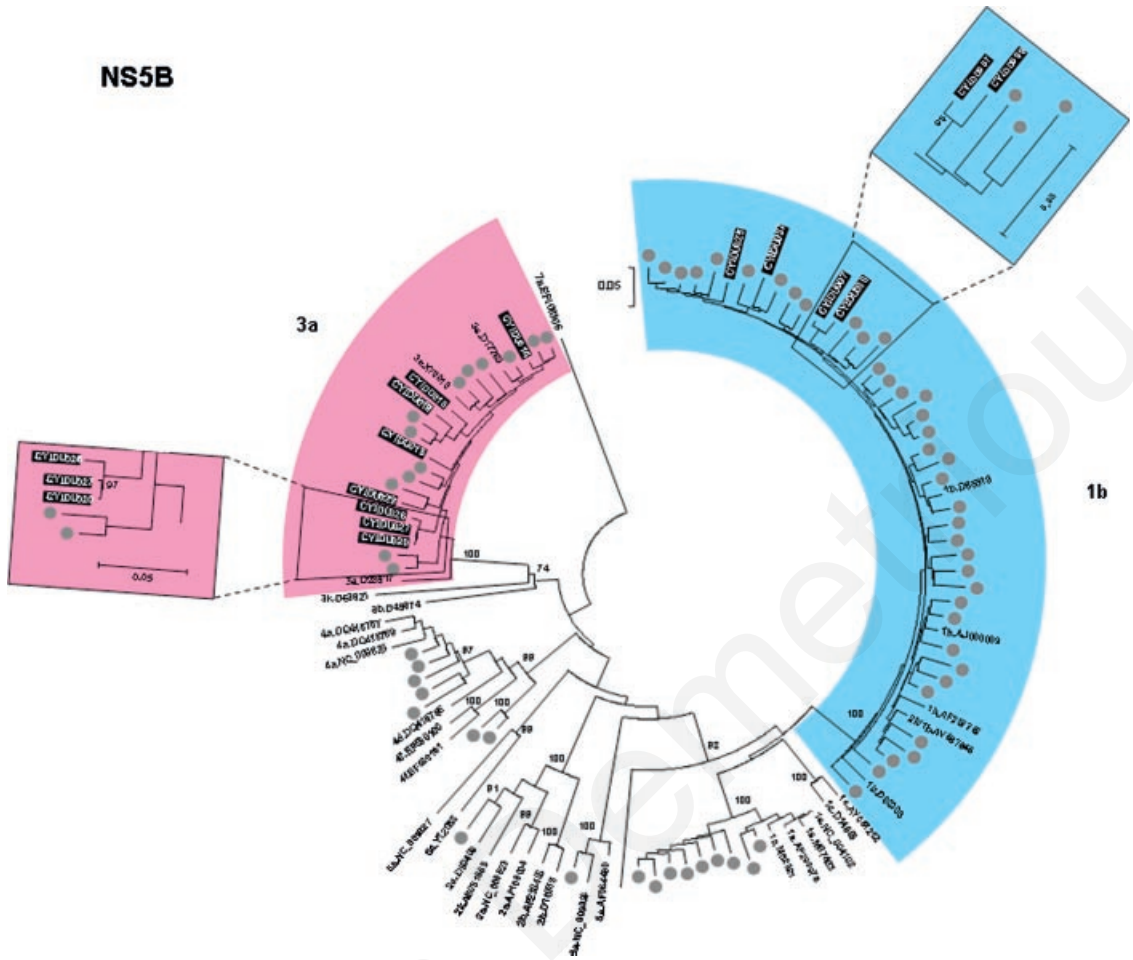

Fig. 3. Neighbor-joining tree of NS5B sequences, constructed as described in the "Materials and Methods" section. The intravenous drug users' strains are indicated with their laboratory codes and highlighted in black with white font. The reference strains are written in black font. The strains of the general population are indicated with gray circles. The subtypes in which the injecting drug users' strains fall are colored: 1b is blue and 3a is pink. The observed clusters of intravenous drug users' strains are indicated with a square and magnified for better viewing. The numbers indicated at the nodes are percentage bootstrap support values for 1,000 replicates. The divergence between any two sequences is obtained by summing the branch length using the scale at the top of the tree.

Laar et al., 2005; Shapatava et al., 2006; Sutton et al., 2008; Tan et al., 2008]. On the other hand, it was surprising to find that no subjects were infected with HIV or HBV. The prevalence of HIV and HBV is known to be lower than that of HCV, as fewer sharing partners are necessary to sustain HCV transmission than for other blood-borne viruses, and indirect drug sharing and preparation practices have also been associated with HCV transmission [Crofts et al., 1999; Alter, 2007], which it is considered applies to this population since the percentage of injecting drug users requesting therapy which was sampled in this study is high (87%).

From studies carried out on intravenous drug users, the main genotypes found to be associated with this transmission route are 1a and 3a [Esteban et al., 2008]. Subtyping of the sequences obtained in the Core-E1 and NS5B region of the HCV genome and, the frequent presence of 3a strains in this population have been confirmed, but subtype 1a, which appears in the general population infected with HCV in Cyprus [Demetriou et al., 2009], was not found in any individuals of this cohort. Interestingly, 43% of the strains were subtype 1b, showing that this subtype has a significant prevalence in intravenous drug users on the island, and possibly illustrates spill-over between different risk groups in this geographical region and the countries where the strains have possibly originated from, that is, eastern and southern Europe. This would have to be confirmed by phylogenetic analysis with HCV sequences from these countries. About 10% of the sequences submitted into the Los Alamos HCV database as ones which have been obtained from intravenous drug users are subtype 1b (http://hcv.lanl.gov), and it has been discovered in other such studies, mostly at a lower prevalence [Mathei et al., 2005; Zhou et al., 2006; Tallo et al., 2007; Peng et al., 2008; Oliveira Mde et al., 2009]. Finally, no 2k/1b strains were found in this group, even though such isolates have been identified on the island and they are known to circulate among intravenous drug users [Kalinina et al., 2002; Demetriou et al., 2009].

The phylogenetic analysis of the sequences resulted in no clustering between the injecting drug users' samples

and those of the general population. Even with such a permissive arbitrary threshold of genetic distance as 0.05, there were no other clusters observed, confirming the polyphyletic nature of HCV infection on the island and illustrating that the HCV epidemic in the selected population runs in parallel to the general population infected with HCV.

Three small clusters were observed within the intravenous drug users group itself, with small genetic distances (<0.05) and high bootstrap values. These clusters are between samples CYIDU06, CYIDU07, and CYIDU33 within genotype 1b, CYIDU24 and CYIDU37 also within 1b, and CYIDU20, CYIDU26, and CYIDU27 within genotype 3a. In the first cluster (CYIDU06, CYIDU07, and CYIDU33), the sequences were obtained from three males between the ages of 24 and 47, all sampled from the same town, two of which from the same center. Two were Cypriot and the other was Greek. They all stated that they had practiced intravenous drug abuse within the past year, and interestingly all had been practicing intravenous drug use for over 10 years. In the second cluster (CYIDU24 and CYIDU37) the samples were taken from a 27-year-old Bulgarian male and a 31-year-old Cypriot male, sampled in different towns. They both mention the use of speedball as the drug used when first practicing injecting drug use, and have both been practicing intravenous drug use for over 13 years. Only one other individual reported this, CYIDU02, however this person tested negative in the HCV antibody testing. In the third cluster (CYIDU20, CYIDU26, and CYIDU27), are samples from three Cypriots, one female and two males aged between 19 and 24. Two of the samples were obtained from users in the same detoxification clinic and the other from another center in another town. The males and females from the same detoxification clinic stated a duration of intravenous drug abuse as less than five years and the other individual did not state how long he had been using. These small clusters indicate the probability of sharing injecting equipment between these users. Also Cyprus is a small island and easy to get around, with a relatively small injecting drug user community. Therefore it is not surprising that strains cluster from users who were sampled in different towns or rehabilitation centers.

The Bayesian analysis showed that there was no clustering between the Cypriot intravenous drug users' sequences and any sequence from the Los Alamos HCV database (http://hcv.lanl.gov). Lack of sequence data in the database impairs the ability to investigate molecular epidemiological and phylogenetic relationships between newly obtained sequences and HCV sequences from anywhere around the world. From these results it cannot be determined where the HCV infections were transmitted, for them to make their way into Cyprus through intravenous drug users. More sequence data from intravenous drug users, especially from eastern and southern European countries, would be required to obtain a better picture of the HCV epidemic and to gain insight on the origin and dynamics of HCV infection on a population level. This point is the most severe limitation to this study and to all work on phylogenetic analysis of HCV strains.

In this population of highly experienced intravenous drug users the only significant factor predicting HCV was the duration of injection drug use, a factor known to be a predictor of higher HCV prevalence [Shepard et al., 2005; Hagan et al., 2007; Sweeting et al., 2009]. Also, having served a prison sentence appears to be reported more frequently in HCV-positive individuals, highlighting the occurrence of unsafe injecting practices in correctional facilities. Interestingly, there were no significant differences between the HCV-positive and negative groups with regards to having shared injecting equipment or having easy access to clean syringes, or other stated risk behaviors. All subjects stated that they had practiced unprotected sex, but only two stated MSM behavior, given the recent reports regarding MSM being a route of transmission [van de Laar et al., 2007].

This study is the first description of the molecular epidemiology of HCV infection among intravenous drug users in Cyprus and provides confirmation of the polyphyletic nature of the epidemic, already seen in the general population on the island infected with HCV. This is seen by the limited clusters appearing within the sequences obtained from injecting drug users and none with those obtained from the general population. Also, there appears to be a circulation of only subtypes 1b and 3a among intravenous drug users seeking therapy in Cyprus, and no indication of subtype 1a. Furthermore, the results showed that the duration of intravenous drug abuse was a main factor to acquiring HCV infection compared to other risk factors in this cohort. The prevalence of HCV infection among the studied population was 50%, comparable to the rest of the world but also significant enough to raise an alarm for the requirement of harm reduction strategies to be applied for the protection of intravenous drug users, especially considering the high influx of tourists and political refugees on the island. Overall, this study reinforces the idea of multiple points of introduction of HCV into Cyprus and also the risk of widespread transmission of these strains. Further investigation would be necessary with a more extended sampling group and by sampling through snowballing instead of through detoxification services, to obtain a more complete picture of the epidemiology of HCV among intravenous drug users on the island and rule out the presence of more genotypes and subtypes in this population.

ACKNOWLEDGMENTS

The authors thank all participating subjects and the staff at the drug rehabilitation centers for assisting in sample collection. This work has obtained bioethical approval from the Cyprus National Bioethics Committee (EEBK/E∏/2008/15). The work was supported by a grant from the Cyprus Anti-drugs Council (item 633 of the Budget for Research Programmes 2008) and the

Cyprus Research Promotion Foundation, grant number ΕΝΙΣΧ/0506/34.

REFERENCES

Alter MJ. 2007. Epidemiology of hepatitis C virus infection. World J Gastroenterol 13:2436–2441.

Crofts N, Aitken CK, Kaldor JM. 1999. The force of numbers: Why hepatitis C is spreading among Australian injecting drug users while HIV is not. Med J Aust 170:220–221.

de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, Snoeck J, van Rensburg EJ, Wensing AM, van de Vijver DA, Boucher CA, Camacho R, Vandamme AM. 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21:3797–3800.

Demetriou VL, van de Vijver DA, Kostrikis LG. 2009. Molecular epidemiology of hepatitis C infection in Cyprus: Evidence of polyphyletic infection. J Med Virol 81:238–248.

Esteban JI, Sauleda S, Quer J. 2008. The changing epidemiology of hepatitis C virus infection in Europe. J Hepatol 48:148–162.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Hagan H, Des Jarlais DC, Stern R, Lelutiu-Weinberger C, Scheinmann R, Strauss S, Flom PL. 2007. HCV synthesis project: Preliminary analyses of HCV prevalence in relation to age and duration of injection. Int J Drug Policy 18:341–351.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Kalinina O, Norder H, Mukomolov S, Magnius LO. 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J Virol 76:4034–4043.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

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

Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. 2008. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med 5:e50.

Mathei C, Wollants E, Verbeeck J, Van Ranst M, Robaeys G, Van Damme P, Buntinx F. 2005. Molecular epidemiology of hepatitis C among drug users in Flanders, Belgium: Association of genotype with clinical parameters and with sex- and drug-related risk behaviours. Eur J Clin Microbiol Infect Dis 24:514–522.

Moreau I, Hegarty S, Levis J, Sheehy P, Crosbie O, Kenny-Walsh E, Fanning LJ. 2006. Serendipitous identification of natural intergenotypic recombinants of hepatitis C in Ireland. Virol J 3:95.

Morice Y, Cantaloube JF, Beaucourt S, Barbotte L, De Gendt S, Goncales FL, Butterworth L, Cooksley G, Gish RG, Beaugrand M, Fay F, Fay O, Gonzalez JE, Martins RM, Dhumeaux D, Vanderborght B, Stuyver L, Sablon E, de Lamballerie X, Pawlotsky JM. 2006. Molecular epidemiology of hepatitis C virus subtype 3a in injecting drug users. J Med Virol 78:1296–1303.

Oliveira Mde L, Bastos FI, Telles PR, Hacker Mde A, Oliveira SA, Miguel JC, Yoshida CF. 2009. Epidemiological and genetic analyses of Hepatitis C virus transmission among young/short- and long-term injecting drug users from Rio de Janeiro, Brazil. J Clin Virol 44:200–206.

Peng JS, Wang X, Liu MQ, Zhou DJ, Gong J, Xu HM, Chen JP, Zhu HH, Zhou W, Ho WZ. 2008. Genetic variation of hepatitis C virus in a cohort of injection heroin users in Wuhan, China. Virus Res 135:191–196.

Pybus OG, Cochrane A, Holmes EC, Simmonds P. 2005. The hepatitis C virus epidemic among injecting drug users. Infect Genet Evol 5:131–139.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Shapatava E, Nelson KE, Tsertsvadze T, del Rio C. 2006. Risk behaviors and HIV, hepatitis B, and hepatitis C seroprevalence among injection drug users in Georgia. Drug Alcohol Depend 82:S35–S38.

Shepard CW, Finelli L, Alter MJ. 2005. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis 5:558–567.

Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin IT, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology 42:962–973.

Sutton AJ, Hope VD, Mathei C, Mravcik V, Sebakova H, Vallejo F, Suligoi B, Brugal MT, Ncube F, Wiessing L, Kretzschmar M. 2008. A comparison between the force of infection estimates for blood-borne viruses in injecting drug user populations across the European Union: A modelling study. J Viral Hepat 15:809–816.

Sweeting MJ, Hope VD, Hickman M, Parry JV, Ncube F, Ramsay ME, De Angelis D. 2009. Hepatitis C Infection Among Injecting Drug Users in England and Wales (1992–2006): There and Back Again? Am J Epidemiol 170:352–360.

Sy T, Jamal MM. 2006. Epidemiology of hepatitis C virus (HCV) infection. Int J Med Sci 3:41–46.

Tallo T, Norder H, Tefanova V, Krispin T, Schmidt J, Ilmoja M, Orgulas K, Pruunsild K, Priimagi L, Magnius LO. 2007. Genetic characterization of hepatitis C virus strains in Estonia: Fluctuations in the predominating subtype with time. J Med Virol 79:374–382.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEG A4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599.

Tan Y, Wei QH, Chen LJ, Chan PC, Lai WS, He ML, Kung HF, Lee SS. 2008. Molecular epidemiology of HCV monoinfection and HIV/HCV coinfection in injection drug users in Liuzhou, Southern China. PLoS ONE 3:e3608.

van Asten L, Verhaest I, Lamzira S, Hernandez-Aguado I, Zangerle R, Boufassa F, Rezza G, Broers B, Robertson JR, Brettle RP, McMenamin J, Prins M, Cochrane A, Simmonds P, Coutinho RA, Bruisten S. 2004. Spread of hepatitis C virus among European injection drug users infected with HIV: A phylogenetic analysis. J Infect Dis 189:292–302.

van de Laar TJ, Langendam MW, Bruisten SM, Welp EA, Verhaest I, van Ameijden EJ, Coutinho RA, Prins M. 2005. Changes in risk behavior and dynamics of hepatitis C virus infections among young drug users in Amsterdam, the Netherlands. J Med Virol 77:509–518.

van de Laar TJ, van der Bij AK, Prins M, Bruisten SM, Brinkman K, Ruys TA, van der Meer JT, de Vries HJ, Mulder JW, van Agtmael M, Jurriaans S, Wolthers KC, Coutinho RA. 2007. Increase in HCV incidence among men who have sex with men in Amsterdam most likely caused by sexual transmission. J Infect Dis 196:230–238.

Zhou DX, Tang JW, Chu IM, Cheung JL, Tang NL, Tam JS, Chan PK. 2006. Hepatitis C virus genotype distribution among intravenous drug user and the general population in Hong Kong. J Med Virol 78:574–581.

Journal of Medical Virology 81:238–248 (2009)

Molecular Epidemiology of Hepatitis C Infection in Cyprus: Evidence of Polyphyletic Infection

Victoria L. Demetriou,[1] David A.M.C. van de Vijver,[2] The Cyprus HCV Network,[†] and Leondios G. Kostrikis[1]*

[1]*Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus*
[2]*Department of Virology, Erasmus MC, University Medical Centre Rotterdam, The Netherlands*

The genetic diversity of the hepatitis C virus (HCV) in Cyprus is investigated for the first time in this study. Nucleotide sequence analysis of the CORE-E1 and NS5B regions of the HCV genome was performed on blood plasma samples obtained from 77 HCV patients in Cyprus, collected during 2005–2008. The amplified products were sequenced and compared to reference HCV strains of known genotype and subtype in order to classify the isolates found in this study. Genotype could be determined for all strains, and subtype for all but four isolates. Phylogenetic analysis revealed that 51 patients were genotype 1, of which 38 were subtype 1b, 9 were 1a, and 1 was unclassified, one patient was genotype 2c, 13 were genotype 3a, nine were genotype 4, of which six were subtype 4a, and three were of unclassified subtype, one was genotype 5a, two patients seem to carry a possible 2k/1b recombinant strain, and no genotype 6 strains were found. This study demonstrated a genetic heterogeneity of HCV infection in Cyprus, with five of the six known HCV genotypes on the island, including unclassified isolates in genotypes 1 and 4, and also the apparent introduction of the 2k/1b recombinant strain in intravenous drug users. ***J. Med. Virol. 81:238–248, 2009.*** © 2008 Wiley-Liss, Inc.

KEY WORDS: HCV genotypes; NS5B region; CORE-E1 region; phylogenetic analysis

INTRODUCTION

Since its identification in 1989 [Choo et al., 1989], hepatitis C has been recognized as a major public health problem infecting nearly 170 million people around the world [WHO, 1999]. It is a life-shortening disease associated with complex and expensive morbidity and decreased quality of life, being a major contributor to liver cirrhosis and hepatocellular carcinoma. In 15–20% of acute HCV infections the patient recovers spontaneously, but in the large majority of cases the disease runs a chronic course and can even cause hepatocellular carcinoma [Seeff, 2002]. The most important route of HCV transmission is through exposure to infected blood and until the introduction of diagnostic screening in 1991, the virus was transmitted mainly through blood, blood products, hemodialysis, and organ transplantation [Memon and Memon, 2002]. HCV infection in the Western world now occurs primarily by parenteral exposure, the most common mode of transmission being intravenous drug use through sharing of needles or other injecting equipment. However, in developing countries unsafe therapeutic injection practices, inadequate disinfection practices, non-sterile medical and dental procedures, and unscreened blood transfusions may still account for significant HCV transmission and serve as a bridge to the general population. Although certain recent reports also link HCV transmission with sexual behavior [van de Laar et al., 2007; Richardson et al., 2008], this has been a controversial association for many years.

[†]The Cyprus HCV Network investigators: S. Chimonides, Nicosia General Hospital, Nicosia. A. Evgeniou, Evgeniou Clinic, Larnaca; E. Hadjigeorgiou-Vounou, Limassol General Hospital, Limassol; M. Koliou-Mazeri, Hospital of Archbishop Makarios III, Nicosia; P. Papakyriakou, Paphos General Hospital, Paphos; L. Petsas, Larnaca General Hospital, Larnaca; G. Potamitis, Gastroenterology Clinic, Nicosia.

Grant sponsor: European Commission; Grant numbers: FP6-014822, QLK2-CT-2001-01344, LSHP-CT-2006-518211; Grant sponsor: Cyprus Research Promotion Foundation; Grant number: ΕΝΙΣΧ/0506/34; Grant sponsor: University of Cyprus (to L.G.K.); Grant numbers: 8037-3/312-25004, 837-25011; Grant sponsor: Birch Biomedical Research LLC (to L.G.K.); Grant number: 3416-25017.

*Correspondence to: Leondios G. Kostrikis, Laboratory of Biotechnology and Molecular Virology, Department of Biological Sciences, University of Cyprus, 75 Kallipoleos Avenue, PO Box 20537, 1678 Nicosia, Cyprus. E-mail: lkostrik@ucy.ac.cy

Accepted 10 September 2008

DOI 10.1002/jmv.21370

Published online in Wiley InterScience (www.interscience.wiley.com)

The hepatitis C virus (HCV) is a small, enveloped, single-stranded, positive-sense RNA virus belonging to the genus *Hepacivirus* of the family Flaviviridae, consisting of a genome of approximately 9,500 nucleotides. Studies of the nucleotide sequences of HCV variants from different individuals and different geographical regions have revealed a high degree of genetic heterogeneity of the HCV genome, and this has led to a proposed consensus of six genotypes and numerous closely related subtypes based on sequence variation in the 5′ noncoding region (5′NCR), CORE, E1 and NS5B regions [Simmonds et al., 1993, 2005; Robertson et al., 1998; Simmonds, 1999]. At the level of full genomes, HCV exhibits around 30% variation between genotypes and 20–25% variation in subtypes of the same genotype. The genotype is presently a major factor in both the choice of treatment and prognosis [Zeuzem, 2004].

Genotype distribution differs by geographic region and by year and mode of transmission [Zein, 2000; Schroter et al., 2004]. Globalization, however, is changing the face of HCV epidemiology radically. Knowledge of genotype distribution in different parts of the world may help clarify the epidemiology and evolution of HCV, and has proven a useful tool for identifying risk groups and distinguishing different routes of transmission [Nakano et al., 2004]. Certain modes of transmission are associated with HCV subtype infections, suggesting separate HCV epidemics, but spill-over between different risk groups underlines the value of molecular epidemiological studies to gain insight into the origin and dynamics of HCV infections on a population level. In western Europe, HCV subtypes 1a and 3a predominate among intravenous drug users [Cochrane et al., 2002; van Asten et al., 2004] whereas subtypes 1b and genotype 2 are associated mainly with contaminated blood transfusions and other types of nosocomial transmission, especially in older patients. Geographically, genotypes 1, 2, and 3 are found globally, while 4–6 have a more restricted pattern. Genotype 4 is found mainly in North Africa and especially Egypt, but has recently been spreading to Europe largely through intravenous drug users with a high incidence in Greece [Savvas et al., 2005; Katsoulidou et al., 2006; Kamal and Nasser, 2008]. Genotype 5 is restricted primarily to South Africa [Chamberlain et al., 1997], but has also recently been found in West Flanders, Belgium [Verbeeck et al., 2006], central France [Henquell et al., 2004], and Syria [Antaki et al., 2008], albeit in much smaller numbers. HCV genotype 6 is found in Southeast Asia [Huy and Abe, 2004].

In the geographical area close to Cyprus, which is the eastern Mediterranean region, HCV genotypes are not distributed uniformly. In western Turkey, genotype 1b is the most prevalent [Altuglu et al., 2007]. Also, a study done in northern Cyprus with civilians, Turkish soldiers, and Northern Cyprus soldiers revealed genotype 1b as the most prevalent (92.4%) [Altindis et al., 2006]. In Egypt, the incidence of HCV infection is significantly higher than other countries worldwide and most cases are infected with subtype 4a [Abdel-Hamid et al., 2007]. Studies carried out with patients in the Middle East revealed a predominance of HCV genotypes 4 and 1 [Watson et al., 1999; Ramia and Eid-Fares, 2006]. In Greece, genotype 1 is the most prevalent (46.9%), followed by genotype 3 (28.1%), 4 (13.2%), 2 (6.9%), and 5 (0.4%) [Katsoulidou et al., 2006]. This pattern of diversity is much more similar to the results presented in this study.

Among injected drug users, the HCV subtypes most prevalent are 3a and 1a, with both subtypes showing an exponential population growth during the 20th century [Pybus et al., 2005]. Genotype 3a, which originates in Asia, has been associated significantly with transmission through intravenous drug use in industrialized countries [Pawlotsky et al., 1995; McCaw et al., 1997; Bourliere et al., 2002]. It is prevalent mainly in North and South America, Europe, and Australia where practicing intravenous drug abuse is common, and seems to have been transmitted from a common origin through a unique worldwide epidemic that spread rapidly among drug users [Pybus et al., 2005; Morice et al., 2006]. Genotype 4 (mainly 4d) is also becoming increasingly prevalent in populations of intravenous drug users, especially in southern Europe, and its introduction into the European intravenous drug user population seems to be more recent than that of 1a and 3a [van Asten et al., 2004; Chlabicz et al., 2008; Kamal and Nasser, 2008].

Until 2001, HCV was thought to evolve in a clonal manner, with diversity generated through the accumulation of mutations. However, homologous recombination has been demonstrated between different genotypes or different subtypes of a genotype. A 2k/1b recombinant was found in St. Petersburg, Russia [Kalinina et al., 2002], and seems to be spreading among intravenous drug users in Russia. It has also been found in Ireland [Moreau et al., 2006], in Estonia [Tallo et al., 2007] and among intravenous drug users in Uzbekistan [Kurbanov et al., 2007]. Other natural recombinants of the virus found in certain parts of the world are a 2i/6p recombinant in Vietnam [Nopporpanth et al., 2006], a 2b/1b recombinant in the Philippines [Kageyama et al., 2006], a 1b/1a recombinant in Peru [Colina et al., 2004] and one between genotypes 2 and 5 in southwest France [Legrand-Abravanel et al., 2007]. HCV recombination break points have been located mainly in the non-structural proteins, but an intratypic recombinant with a break point in the structural region has also been identified [Cristina and Colina, 2006].

Cyprus is a small island with a population of approximately 800,000 and a large annual influx of foreigners mainly from tourism but also as political refugees. The molecular epidemiology of hepatitis C infection in Cyprus has never before been studied. The HCV genotype distribution on the island is presented here for the first time, revealing high genetic heterogeneity, multiple points of introduction, and the existence of possible recombinant strains.

MATERIALS AND METHODS

Patients and Samples

From 2005 to 2008 blood samples were obtained from 107 consenting chronically infected HCV patients aged 18–84 attending private clinics and public hospitals in the Nicosia, Larnaca, Limassol, and Paphos districts. All patients were tested positive for HCV antibodies by a second-generation immunoassay (INNO-LiPA), and for HCV RNA by diagnostic reverse transcription-polymerase chain reaction (RT-PCR; COBAS Amplicor, Roche Diagnostics, Branchburg, NJ). All samples were investigated by sequencing the CORE-E1 region and the NS5B region of the HCV genome.

RNA Extraction and RT-PCR

Blood was collected from the patients in BD Vacutainer® PPT™ (Becton Dickinson and Co., Franklin Lakes, NJ) tubes and the plasma was isolated after centrifugation at 1,100 RCF (relative centrifugal force) for 10 min in an Eppendorf Centrifuge 5810 R (Eppendorf). Viral RNA was extracted from 200 μl plasma using the QIAmp® UltraSens® Virus kit (Qiagen, Venlo, The Netherlands) and 15 μl of the RNA was used in a one-step RT-PCR using Superscript™ III One-Step RT-PCR Platinum Taq HiFi (Invitrogen, Carlsbad, CA), following a heat-shock step at 70°C for 20 sec to denature the RNA secondary structure. The RT-PCR was performed in a 50 μl reaction with 20 pmol each of the outer sense and antisense degenerate primers derived from the CORE-E1 and NS5B regions of the HCV genome, designed to amplify all HCV genotypes (see Table I). A nested PCR was performed using 3 μl of the RT-PCR product with 40 pmol each of the inner PCR primers (Table I), using Platinum® PCR SuperMix (Invitrogen) in a 50 μl reaction. PCR amplification was confirmed by visualization with ethidium bromide staining of a 2% agarose gel.

CORE-E1 and NS5B Sequencing and Phylogenetic Analysis

Cycle sequencing PCR was performed on the amplicons in both directions using the inner forward and reverse amplification primers for each region (Table I) by means of the BigDye® Terminator system v3.1 (Applied Biosystems, Foster City, CA). The products were purified using the DyeEx spin kits (Qiagen) and sequenced directly on the ABI 3300 Genetic Analyser (Applied Biosystems). The resulting readings were analyzed with the Sequencing Analysis Software v5.2 (Applied Biosystems). The obtained nucleotide sequences of both the CORE-E1 region (77 sequences, 417 bp, positions 867–1283) and NS5B region (70 sequences, 405 bp, positions 8277–8681) were aligned with the reference sequence of the H77 strain using the CLUSTALX 1.83 alignment software [Thompson et al., 1997]. Subtyping the sequence of each region was performed using Oxford HCV Subtyping Tool v1.0 [de Oliveira et al., 2005], after which the aligned sequences were compared to reference strains of known subtypes derived from the Los Alamos database [Kuiken et al., 2005] using the neighbor-joining method [Saitou and Nei, 1987] in MEGA version 4 [Tamura et al., 2007]. Pair-wise distance matrices were generated using the Kimura [1980] two-parameter distance estimation approach. The reliability of the phylogenetic clustering was evaluated using bootstrap analysis with 1,000 replicates [Felsenstein, 1985]. Bootstrap values above 70 were considered sufficient for subtype assignment.

NS5B Phylogenetic Analysis of 3a Strains

Further phylogenetic analysis was performed on the isolated genotype 3a strains found in this study by constructing a tree from the NS5B sequences of these strains, and 50 intravenous drug use-related NS5B sequences from published work about strains from intravenous drug users [Kalinina et al., 2001; Cochrane et al., 2002; Morice et al., 2006] and other subtype 3a sequences with intravenous drug use as the stated source of infection retrieved from the HCV sequence database. The analysis was done using the neighbor-joining method [Saitou and Nei, 1987] in MEGA version 4 [Tamura et al., 2007]. Pair-wise distance matrices were generated using the Kimura [1980] two-parameter distance estimation approach. The reliability of the phylogenetic clustering was evaluated using bootstrap analysis with 1,000 replicates [Felsenstein, 1985].

Reference Sequences

The GenBank accession numbers for reference sequences used in phylogenetic analysis of the CORE-E1 region are: AB031663, AF064490, AF165045, AF169004, AF238486, AF271822, AF271876, AF271878, AF271886, AF290978, AJ000009, AY051292, AY434107, AY434119, AY434122, AY434128, AY434131, AY434134, AY434146, AY434149, AY434158, AY587845, AY706996, AY706999, AY754623, AY767506, AY767956, AY894540, AY894555, D10988, D14853, D28917, D43678, D50409, D63821, D90208, DQ418786, DQ418787, DQ418789, E10839, EF115767, EF115770, EF115798, EF115882, EF115883, EF115898, EF115900, EF115902, EF115906, EF115908, EF115915, EF115916, EF115923, EF589160, EF589161, L29589, L29609, L29610, L29620, L38350, L39282, L39310, NC_004102, NC_009823, NC_009824, X76414, Y11604, Y12083, Y13184. The GenBank accession numbers of the sequences used as references for phylogenetic analysis of the NS5B region are: AB031663, AF037235, AF037237, AF064490, AF165045, AF169004, AF238486, AF271799, AF290978, AJ000009, AY051292, AY265429, AY265435, AY434106, AY434108, AY434120, AY434123, AY434126, AY434132, AY434147, AY434157, AY548714, AY548717, AY548731, AY548736, AY587845, AY632098, AY632126, AY632144, AY632237, AY685046, AY743124, AY743160, AY743171, AY743182, AY743204, AY743208, AY743212, AY743213, AY754624, AY894553, D10988, D14853, D28917, D50409, D63821, D90208, DQ418786, DQ418787, DQ418789, DQ911240, E10839, EF115983,

TABLE I. RT-PCR Primers for CORE-E1 and NS5B Amplification and Sequencing

Name	Primer set	Polarity	Sequence[a]	Position[b]
CORE-E1				
735	Outer	Sense	5′-GACCTCATGGGGTACATYCCBSTCGTHGG-3′	735–763
1324	Outer	Antisense	5′-GGBGACCARTTYAKCATCATRTCCCAWGCC-3′	1,295–1,324
834	Inner	Sense	5′-GCAACAGGGAATYTDCCYGGTTGCTCYTTYTC-3′	834–865
1318	Inner	Antisense	5′-CAGTTCATCATCATRTCCCAWGCCATNCGRTGDCC-3′	1,284–1,318
NS5B				
8172	Outer	Sense	5′-TAYGGRTTCCARTACTCNCCHGVRCAGCGGGT-3′	8,172–8,203
8821	Outer	Antisense	5′-GARTTGACWGGRGWGTGTCKDRCTGTYTCCCA-3′	8,790–8,821
8244	Inner	Sense	5′-ATGGGBTTYKCRTATGAYACCCGHTGYTTTGA-3′	8,244–8,275
8713	Inner	Antisense	5′-GABACRTTKGAGGARCADGATGTTATNARCTC-3′	8,682–8,713

[a]Degenerate positions are shown with their IUB Base Codes (R: A or G; W: A or T; S: G or C; K: G or T; Y: C or T; B: C, G or T; D: A, G or T; H: A, C or T; V: A, C or G; N: A, C, G or T).
[b]Position numbering according to strain H77 (GenBank Acc. No NC_004102), genotype 1a.

EF115994, EF116013, EF116021, EF116118, EF116121, EF116125, EF116137, EF116138, EF116141, EF589160, EF589161, L29611, L29618, L38371, L48496, NC_004102, NC_009823, NC_009824, Y11604, Y12083, Y13184. The GenBank accession numbers of the reference sequences used in the subtype 3a tree for intravenous drug user strains are: AB327108, AB327110, AB327111, AB327112, AB327113, AB327114, AB327115, AF388439, AF388443, AF388447, AF388450, AF388451, AF388452, AF388455, AF388464, AF388466, AF388467, AF388469, AF388475, AF388476, AF388509, AF516369, AJ867081, AJ867088, AJ867093, AJ867098, AJ867101, AJ867105, AJ867106, AJ867162, AY100024, AY100031, AY100037, AY100045, AY100047, AY100051, AY100052, AY100055, AY100074, AY100077, AY100079, AY100081, AY100083, AY100084, AY100090, AY100093, AY100095, AY100107, AY100109, AY100111 and the tree was rooted with non-a genotype 3 strains E10839 and D63821.

Statistical Analysis

To test for a statistically significant correlation between the PCR results and demographic and clinical variables from the samples, the χ^2-test for categorical variables and the t-test for continuous variables were used.

Nucleotide Sequence Accession Numbers

GenBank accession numbers for the sequences obtained in this study are EU684591–EU684660 for the NS5B sequences and EU684661–EU684737 for the CORE-E1 sequences.

RESULTS

Clinical and Epidemiological Features of Study Subjects

The study group consisted of 107 HCV seropositive patients between the ages of 18 and 84, from two private gastroenterology clinics and public hospitals in Cyprus. The epidemiological features of the study subjects varied, as 53 (51.5%) patients were Cypriots and the rest were various other nationalities, 36 (35%) being from countries of the former Soviet Union (Russia, Georgia, Moldova, and Ukraine). Thirty-nine patients (36.4%) had a history of transfusion with blood products, 4 (3.7%) stated they were intravenous drug users, 10 (9.3%) traced infection to dental or surgical procedures, and 51 (47.7%) did not know the source of infection (see Table II for the demographic details).

Viral RNA Extraction From Plasma and RT-PCR

Seventy-seven samples were PCR-positive for CORE-E1 and 70 for NS5B, presenting 73.8% and 71.0% PCR success rates, respectively. Statistical analysis to determine the association between PCR result and whether the patients were on therapy revealed a more frequent negative PCR result for patients on therapy with P-values <0.016. Fifty-six patients (52.3%) were on interferon therapy when blood was taken, and 51 patients (47.7%) were taking ribavirin. Of all patients on treatment, 49 were on interferon–ribavirin combination therapy. Forty-nine patients (45.8%) were not on therapy at the time blood was taken. Considering the two drugs separately, from the patients on interferon therapy, 35 exhibited PCR-positive results and 21 showed negative results, compared to 44 positive results and 7 negative results from the patients not on interferon ($P = 0.008$). Of the patients taking ribavirin, 32 had positive PCR results and 19 had negative results compared to 47 positive and 9 negative samples from patients not taking ribavirin ($P = 0.016$). Furthermore, the correlation between PCR result and viral load showed that samples from patients with a detectable viral load (more than 135 copies/ml) exhibited more frequently a positive PCR result than those with undetectable viral load with a P-value of 0.002. From the patients with undetectable viral load at the time blood was taken, 6 had positive PCR results and 13 had negative results, compared to patients with detectable viral load, of which 58 were positive and 21 were negative for PCR ($P = 0.002$; patients with unknown viral load were excluded from this analysis).

TABLE II. Characteristics of the Study Subjects

Characteristics	Patients (N = 107)	
Gender (%)[a]		
Male	53 (50)	
Female	51 (48)	
Age (years) (%)[a]		
Median (IQR)	40 (32–57)	
18–29	21 (20)	
30–39	31 (30)	
40–49	11 (11)	
50–59	22 (21)	
60 and older	19 (18)	
Region of origin (%)[b]		
Cyprus	53 (52)	
Russia	15 (15)	
Georgia	14 (14)	
Greece	5 (5)	
Moldova	4 (4)	
Ukraine	3 (3)	
Egypt	3 (3)	
Other—Europe	4 (4)	
Other—Asia	2 (2)	
Route of transmission (%)		
Blood transfusion	39 (36)	
Surgical procedure	6 (6)	
Dental procedure	4 (4)	
IVDU	4 (4)	
Tattoo	2 (2)	
Sexual transmission	1 (1)	
Other/unknown	51 (48)	
Use of medication (%)		
Interferon/ribavirin	58 (54)	
No medication	49 (46)	
Plasma HCV-RNA (log copies/ml)		
Median (IQR)	5.5 (2.5-6.2)	
	CORE-E1	NS5B
RT-PCR (%)		
Positive	79 (74)	76 (71)
Negative	28 (26)	31 (29)
Genotype (%)		
1	51 (48)	48 (45)
2	3 (3)	1 (1)
3	13 (12)	13 (12)
4	9 (8)	7 (7)
5	1 (1)	1 (1)
Unknown	30 (28)	37 (35)

IQR, interquartile range.
[a]Information available for 104 patients.
[b]Information available for 103 patients.

CORE-E1 and NS5B Sequencing and Phylogenetic Analyses

The neighbor-joining trees for the CORE-E1 and NS5B regions are seen in Figure 1. HCV genotype 1 was the most frequent (47.7% in the CORE-E1 region), followed by genotypes 3, 4, 2, and 5 (12.1%, 8.4%, 2.8%, and 0.9%, respectively in the CORE-E1 region). There is concordance between the trees for all samples that were positive for both regions, except for two strains from Georgian patients (designated with an asterisk in Fig. 1). These strains were identified as subtype 2k in the CORE-E1 region and as 1b in the NS5B region. From these results, the strains appear to be 2k/1b recombinants, as they cluster together with the St. Petersburg 2k/1b recombinant strain in both trees. However, further clonal analysis is required to confirm this.

For genotype 1, 36 sequences recovered in this study classify with subtype 1b. Two additional strains also classify as subtype 1b in the CORE-E1 region, but were PCR-negative in the NS5B region and could not be analyzed. Twenty of these strains are from patients originating from Cyprus, six from Georgia, five from Russia, three from Moldova, and one each from Greece, Ukraine, and Romania. One strain in particular, from a Greek patient, has a 3-nucleotide insertion in the CORE-E1 sequence between positions 1044 and 1045 (numbers according to strain H77).

Nine strains classify within the 1a subtype in both regions with bootstrap values higher than 90. Three further isolates also classify as subtype 1a in the CORE-E1 region but were PCR-negative in the NS5B region. Of the nine strains, six are from Cypriot patients, two from Georgian patients and one each from a British, an Iranian, an Italian, and a Greek patient. One strain from a Cypriot patient does not cluster with any specific genotype 1 subtype, having used reference sequences from all available assigned genotype 1 subtypes in this genomic region in the phylogenetic trees. This is also the case for the NS5B tree, where all available assigned subtypes in this region were also included in the phylogenetic analysis. To investigate this strain further, BLAST was performed using the HCV BLAST tool on the HCV Sequence database website, recovering the 100 closest matches. These were downloaded into a FASTA file along with the Cypriot strain and a phylogenetic tree was constructed to explore the relationship with the closest sequences available in the database (data not shown). The strain in question did not cluster with any of the sequences from the database. This sequence is thus labeled unclassified, and due to lack of data, could not be assigned as a new subtype.

For genotype 2, one strain from a Cypriot patient clusters in subtype 2c in the CORE-E1 tree and NS5B tree with bootstrap values of 98 and 83, respectively. Also the two putative recombinant strains mentioned

Fig. 1. Neighbor-joining phylogenetic trees for the 77 CORE-E1 sequences (400 nucleotides, corresponding to positions 884–1283; left) and the 70 NS5B sequences (328 nucleotides, corresponding to positions 8277–8604; right) of HCV strains obtained from patients in Cyprus, based on the Kimura two-parameter method for estimation of genetic distance. Trees were constructed using 35 representative reference sequences from 6 known subtypes (1 through 6) and 1 recombinant strain (2k/1b) taken from the HCV sequence database of Los Alamos National Laboratory [Kuiken et al., 2005]. The sequences determined in the study are color-coded, with colors corresponding to the patients' country of origin: Cyprus (yellow), Russia (pink), Georgia (blue), Moldova (light green), Greece (gray), Egypt (lilac), Bulgaria (beige), Iran (light blue), Italy (brown), Pakistan (purple), Romania (black), Ukraine (green), United Kingdom (red). The asterisks indicate putative 2k/1b recombinant strains. The divergence between any two sequences is obtained by summing the branch length, using the scale at the lower left of each tree. The numbers indicated at genotype and subtype-determining nodes are percentage bootstrap support for 1,000 replicates. The brackets on the right side of the trees indicate the determined subtypes as described in Results.

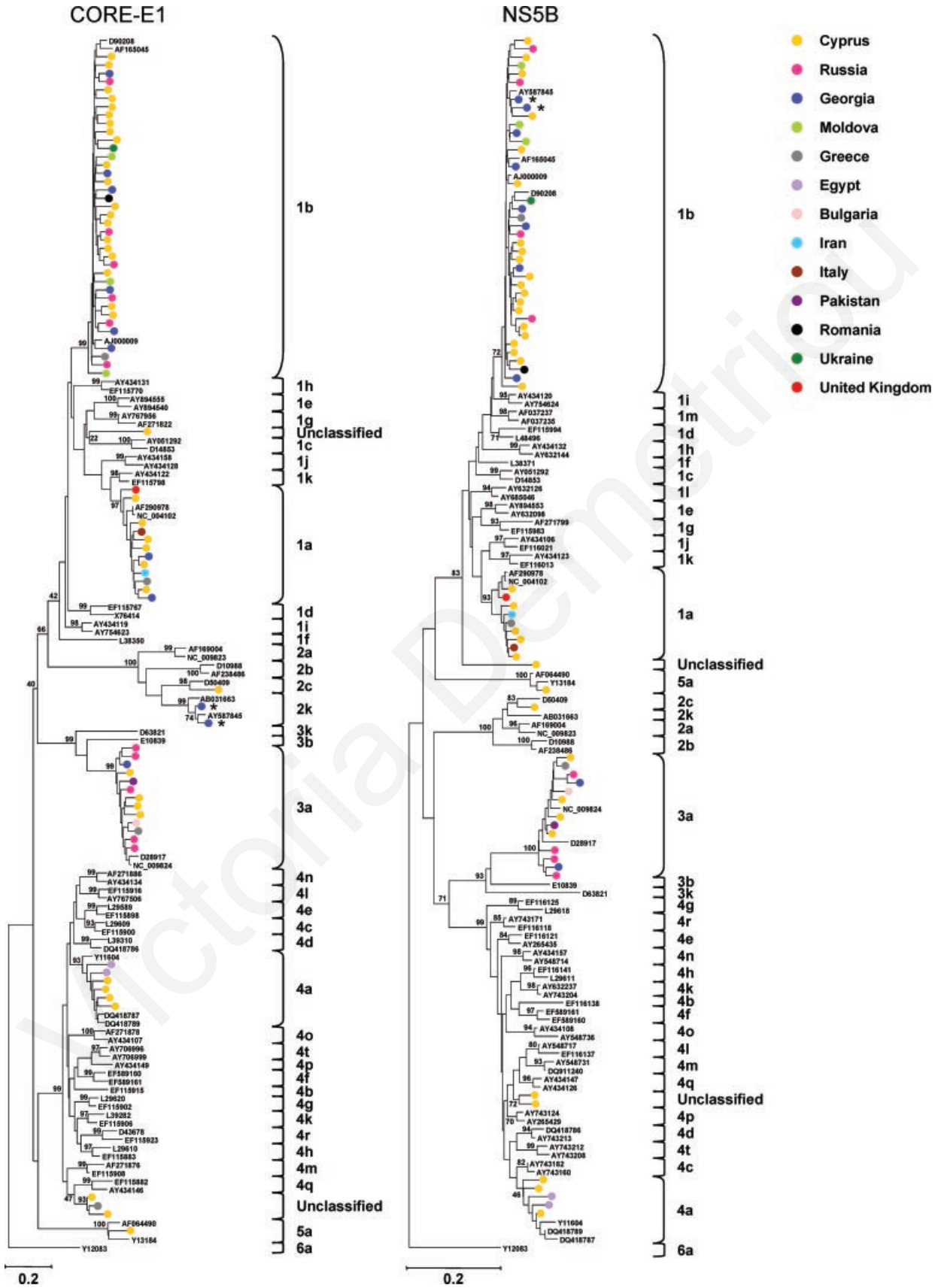
CORE-E1
NS5B
Cyprus
Russia
Georgia
Moldova
Greece
Egypt
Bulgaria
Iran
Italy
Pakistan
Romania
Ukraine
United Kingdom
1b
1h
1e
1g
Unclassified
1c
1j
1k
1a
1d
1i
1f
2a
2b
2c
2k
3k
3b
3a
4n
4l
4e
4c
4d
4a
4o
4t
4p
4f
4b
4g
4k
4r
4h
4m
4q
Unclassified
5a
6a
1b
1i
1m
1d
1h
1f
1c
1l
1e
1g
1j
1k
1a
Unclassified
5a
2c
2k
2a
2b
3a
3b
3k
4g
4r
4e
4n
4h
4k
4b
4f
4o
4l
4m
4q
Unclassified
4p
4d
4t
4c
4a
6a
0.2
0.2

Fig. 1.

earlier cluster with the 2k subtype in the CORE-E1 tree with a bootstrap of 99, but with the 1b subtype in the NS5B region.

Analysis of both genomic regions revealed 13 strains for genotype 3, all of which belong to subtype 3a, clustering with the 3a reference strains with bootstrap values higher than 95 in both trees. These strains were isolated from four Cypriot patients, four Russian patients, two Georgian patients, and one patient each from Greece, Pakistan, and Bulgaria.

For genotype 4, five strains found in this study are assigned as subtype 4a in both regions, plus one additional strain for which only the CORE-E1 region could be sequenced. These strains were isolated from four Cypriot patients and two patients of Egyptian origin. A cluster of a further three strains was found to belong to genotype 4 in the CORE-E1 region but these sequences do not cluster with any particular assigned subtype of genotype 4 and were therefore labeled as unclassified. These strains were isolated from two Cypriot patients and one Greek patient. Two of these strains were seen to behave similarly in the NS5B region, the third being PCR-negative for NS5B, and therefore could not be sequenced in this region. For both trees, sequences from all available subtypes in the corresponding genomic regions were used and this cluster appears to be genetically closest to subtypes 4q, 4m, and 4l. These sequences were further investigated by uploading them into the HCV BLAST tool of the Los Alamos HCV sequence database and retrieving the 100 closest matches. These sequences were analyzed together and a phylogenetic tree was constructed to look for a possible relationship between the strains found in this study and any other similar sequences in the HCV database (data not shown). This was done for the sequences in both regions. In the CORE-E1 region, the unassigned Cypriot strains cluster with seven other sequences from the database, four from subjects of African origin from a Canadian submission (Acc. No. EF115885, EF115899, EF115905, and EF115910) [Murphy et al., 2007] and three from patients of unknown epidemiological and demographic details from a UK submission (Acc. No. AY766949, AY767036, and AY767953), which are also genotype 4 of unassigned subtype. The sequences do not cluster with sequences of assigned subtype and are genetically closest to the subtype 4q group. A similar pattern is revealed in the NS5B tree, where the two Cypriot sequences cluster with the same four Canadian sequences (Acc. No. EF116108, EF116122, EF116128, and EF116133), apart from assigned subtypes and genetically closest to subtype 4q. They cluster with no other isolates and the British sequences were not sequenced in the NS5B region.

Lastly, one genotype 5 strain was also identified in a Cypriot patient and is classified as subtype 5a on both trees with corresponding bootstrap values of 100 in both cases.

No genotype 6 strains were found in this population study.

NS5B Phylogenetic Analysis of Subtype 3a Strains

To further assess the relationship of the genotype 3a sequences discovered in this study with intravenous drug use, these sequences were examined in a new dataset context. A new tree was constructed using the NS5B region of only the 3a strains from this study and 50 3a NS5B sequences derived from publications that studied sequences from intravenous drug users from various countries [Kalinina et al., 2001; Cochrane et al., 2002; Morice et al., 2006] and also from a search for 3a strains from intravenous drug users in the HCV Sequence Database (Fig. 2). As found previously [Morice et al., 2006], the geographical areas of origin of the subjects do not seem to harbor HCV-3a populations distinct from each other. The tree shows no clearly defined subclade of subtype 3a strains isolated from intravenous drug users from different geographical areas; it can therefore not be verified that the HCV-3a strains found in this study are introduced by intravenous drug use in specific geographical regions.

DISCUSSION

In this study, viral RNA extraction from blood plasma, RT-PCR and nucleotide sequence analysis of the CORE-E1 region and NS5B region were used successfully to genotype HCV strains and determine the genetic heterogeneity in a sample study group of patients in Cyprus. The RT-PCR assay design is considered successful as over 70% of samples were PCR-positive in both regions and those that were not belonged mainly to patients receiving therapy and/or with low viral load. The patients investigated were epidemiologically diverse, showing a wide range of ages, countries of origin, and routes of transmission for HCV. Most patients were either from Cyprus or from countries of the former Soviet Union, but the group also included patients from other countries, including Greece, Britain, Pakistan, Italy, Egypt, Iran, Romania, and Bulgaria. The route of transmission for a large percentage of patients was unknown, mainly because this virus can be carried for many years without diagnosis and because safety measures for the prevention of HCV transmission

Fig. 2. Neighbor-joining phylogenetic trees for the 13 NS5B sequences (240 nucleotides, corresponding to positions 8316–8555) of HCV subtype 3a strains obtained from patients in Cyprus, based on the Kimura two-parameter method for the estimation of the genetic distance. Trees were constructed using 50 intravenous drug use-related subtype 3a sequences taken from published studies [Kalinina et al., 2001; Cochrane et al., 2002; Morice et al., 2006] and from searches in the HCV sequence database of Los Alamos National Laboratory [Kuiken et al., 2005]. The tree is rooted with two non-a genotype three reference sequences. The reference sequences are color-coded, with the colors corresponding to the sampling country: Australia (yellow), France (blue), Russia (pink), Uzbekistan (brown), Brazil (gray), United Kingdom (dark green) and USA (purple). The Cypriot strains are indicated with gray circles. The numbers indicated at the nodes are percentage bootstrap support for 1,000 replicates. The divergence between any two sequences is obtained by summing the branch length, using the scale at the lower left of the tree.

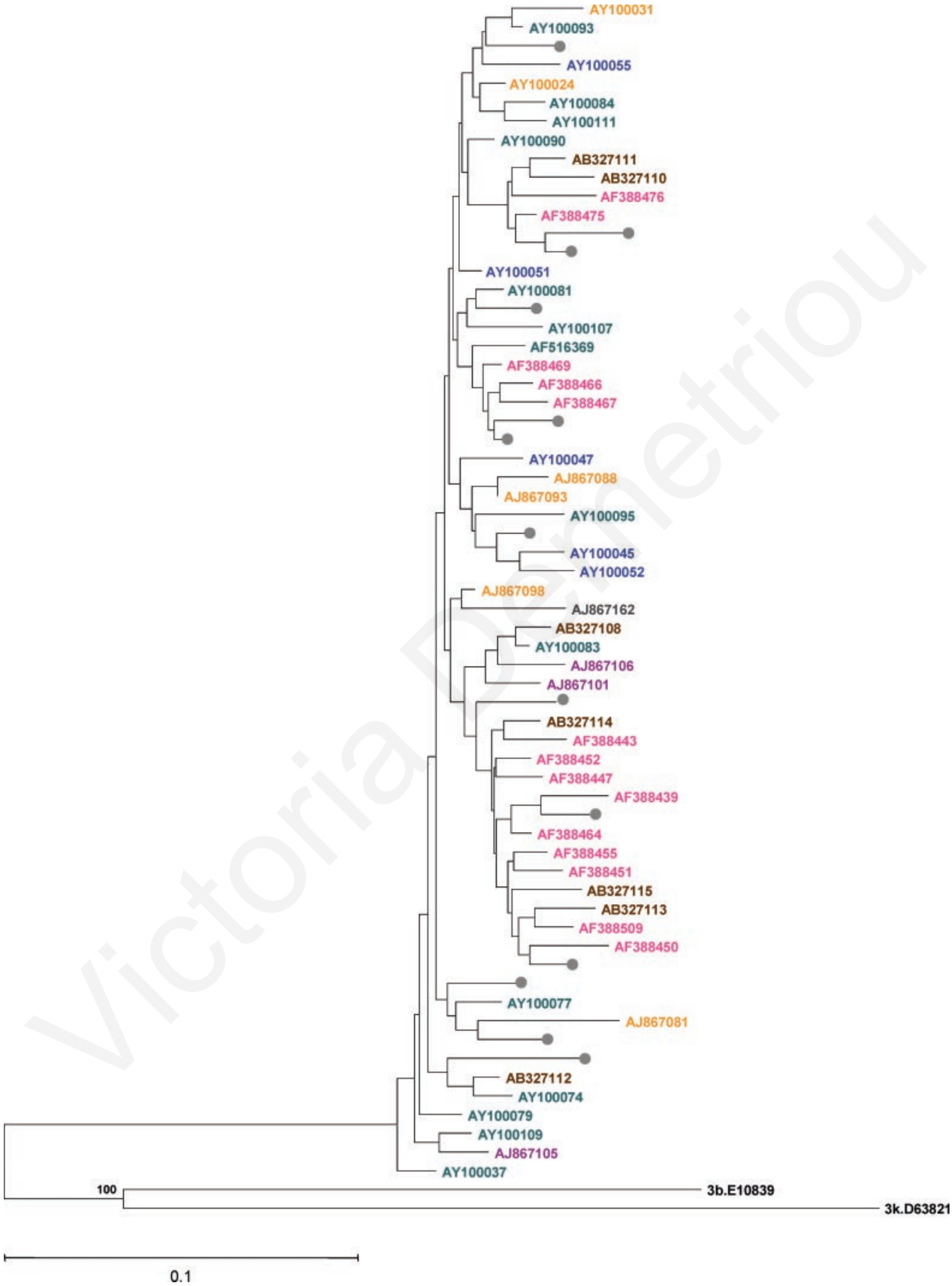
AY100031
AY100093
AY100055
AY100024
AY100084
AY100111
AY100090
AB327111
AB327110
AF388476
AF388475
AY100051
AY100081
AY100107
AF516369
AF388469
AF388466
AF388467
AY100047
AJ867088
AJ867093
AY100095
AY100045
AY100052
AJ867098
AJ867162
AB327108
AY100083
AJ867106
AJ867101
AB327114
AF388443
AF388452
AF388447
AF388439
AF388464
AF388455
AF388451
AB327115
AB327113
AF388509
AF388450
AY100077
AJ867081
AB327112
AY100074
AY100079
AY100109
AJ867105
AY100037
100
3b.E10839
3k.D63821
0.1

Fig. 2.

in hospitals and screening of donated blood and blood products were not carried out until the early 1990s. Also, considering the patients' diversity of nationality, it is worth noting that unsafe nosocomial practice still occurs in developing countries.

For genotype 1, 38 variants belonging to subtype 1b and 12 belonging to subtype 1a were identified. It is interesting that one strain, from a Greek patient, has a three nucleotide insertion in the CORE-E1 sequence between positions 1044 and 1045 (numbers according to strain H77), resulting in an addition of one amino acid in the E1 protein sequence. The epidemiological origin or significance of this insertion is not currently known. One genotype 1 variant has been labeled unclassified as it does not cluster with any of the groups of assigned subtypes for which sequences are available in the HCV sequence database. This strain could be a new subtype in genotype 1, but more epidemiologically distinct strains that cluster with this variant are needed to provisionally assign a new subtype; such sequences were neither found in this study nor in the HCV sequence database.

Of the genotype 2 variants isolated in this study, one strain was identified as subtype 2c. This was isolated from a Cypriot patient who was born in Argentina and was infected there by a blood transfusion in the late 1970s. According to Re et al. [2007], 2c is a subtype found in high prevalence in central Argentina, however it is unknown whether this was the case at the time this patient became infected, nor is it known where in Argentina the blood transfusion took place. Two other strains were classified as subtype 2k in the CORE-E1 region but not in the NS5B region and they are considered to be putative 2k/1b recombinants, discussed below.

For genotype 3, 13 variants belonging to subtype 3a were identified from the analysis of both regions. These strains were isolated from four Cypriot patients, four Russian patients, two Georgian patients, one patient from Greece, one from Pakistan, and one from Bulgaria. Because there is an established correlation of genotype 3a with intravenous drug use [Morice et al., 2006], these strains were analyzed further by comparing them to genotype 3a sequences from intravenous drug users available in the database. This analysis yielded no significant evidence for the Cypriot strains clustering with each other or with isolates from intravenous drug users from other countries. Also it has been shown previously that there is a phylogenetic mixing of HCV subtype 3a strains from drug users and non-drug users in various countries, supporting the existence of a unique origin for subtype 3a [Kalinina et al., 2001; Cochrane et al., 2002; Samimi-Rad et al., 2004; Morice et al., 2006]. The observation of no country-specific phylogenetic clustering for strains isolated from intravenous drug users has been made for all genotypes [van Asten et al., 2004].

For genotype 4, nine strains were identified in this study group, six of which were classified into subtype 4a, four from Cypriot patients, and two from patients originating from Egypt, which is the geographical area with a significantly high prevalence of HCV-4a. It is significant to note the origin of the patients with HCV-4a and the fact that this subtype was not isolated from patients of any other ethnic origin in this study group, which includes a high percentage of patients from countries of the former Soviet Union and other countries. This highlights the fact that this subtype is more restricted to a certain geographical radius than genotypes 1, 2, and 3. Of particular interest was the finding of three isolates that do not classify within any known subtype of genotype 4. Three strains in the CORE-E1 region do not cluster with any particular subtype of genotype 4, having used all available assigned genotype 4 subtypes in this region as reference strains in the phylogenetic analysis. Two of these strains are seen to behave similarly in the NS5B region. The third sample was PCR-negative for NS5B, and therefore could not be sequenced in this region. For both trees, sequences from all available subtypes in the corresponding genomic regions were used.

In genotype 5, just one strain was identified and this was classified as subtype 5a, which is a subtype found primarily in South Africa [Chamberlain et al., 1997], but also in west Flanders, Belgium [Verbeeck et al., 2006], central France [Henquell et al., 2004], and Syria [Antaki et al., 2008]. This strain is from a Cypriot patient who had a transfusion with 17 units (17×450 ml) of blood in 1975 in Johannesburg after a serious accident, but was only diagnosed with hepatitis C in 2006.

The strains indicated on the CORE-E1 and NS5B trees with asterisks (Fig. 1) appear to be 2k/1b recombinants and both were isolated from Georgian patients who stated routes of infection as intravenous drug use and sexual transmission. The first identified 2k/1b recombinant found was recovered in St. Petersburg in intravenous drug users [Kalinina et al., 2002] and has since been found only in Estonia, Ireland, and Uzbekistan in intravenous drug users [Moreau et al., 2006; Kurbanov et al., 2007; Tallo et al., 2007]. The putative recombinants found in this study cluster together and with the St. Petersburg strain in both the CORE-E1 and NS5B trees (Fig. 1). These strains are currently being investigated with clonal analysis along the full genome to identify the putative point of recombination (unpublished results).

Overall, in the Cypriot patients of this study group, all HCV genotypes and subtypes reported in this study were found, except the putative 2k/1b recombinant, revealing high genetic diversity. In the patients coming from countries of the former Soviet Union, the HCV strains identified belonged to subtypes 1b, 1a, 3a, and possibly the 2k/1b recombinant strain, corresponding to the HCV subtypes circulating in the eastern Europe [Naoumov, 1999; Kalinina et al., 2001; Kurbanov et al., 2003; Tallo et al., 2007]. In the Egyptian patients, only subtype 4a was found, reflecting the HCV situation in their country [Abdel-Hamid et al., 2007]. For patients from other European countries (Greece, UK, Italy, Romania, and Bulgaria) the genotypes identified were 1a, 1b, 3a, and 4, which are genotypes found commonly in western and southern Europe [Trepo and Pradat, 1999; Ansaldi et al.,

2005; Katsoulidou et al., 2006; Esteban et al., 2008]. Finally, among patients of Asian ethnicity (Iranian and Pakistani), the genotypes discovered were 1a and 3a, respectively, again corresponding to the most prevalent types of HCV in their countries [Samimi-Rad et al., 2004; Idrees, 2008]. It is, however, difficult to make any further epidemiological conclusions, as for many patients the mode of transmission is unknown and, equally important, so is the country of infection.

The genetic diversity of HCV in Cyprus, as shown in this study, is similar to the findings of HCV diversity in Greece [Katsoulidou et al., 2006], and unlike the findings in Turkey, where subtype 1b is predominant [Altindis et al., 2006; Altuglu et al., 2007], Egypt, which has mainly subtype 4a [Abdel-Hamid et al., 2007], or other countries in the Middle East, where genotypes 4 and 1 predominate [Watson et al., 1999; Ramia and Eid-Fares, 2006]. The heterogeneity in HCV genotypes found in Cyprus is probably due to imported strains from repatriated Cypriots, Cypriots traveling abroad and the large tourism industry. The ethnic background of the study group and the finding of the possible 2k/1b recombinant strains also emphasize the impact of immigrants from eastern Europe and increasing use of intravenous drugs in Cyprus on the multiple points of introduction and risk of widespread transmission of HCV strains on the island.

ACKNOWLEDGMENTS

The authors thank the study participants and the staff at the General Hospitals of Nicosia, Larnaca, Limassol, and Paphos, the Archbishop Makarios III Hospital in Nicosia, Evgeniou private clinic in Larnaca and Dr. G. Potamitis private clinic in Nicosia for assisting in sample collection; G. Hatzipavlou and I. Kousiappa for data preparation; and E. Loizidou for helpful discussions.

REFERENCES

Abdel-Hamid M, El-Daly M, Molnegren V, El-Kafrawy S, Abdel-Latif S, Esmat G, Strickland GT, Loffredo C, Albert J, Widell A. 2007. Genetic diversity in hepatitis C virus in Egypt and possible association with hepatocellular carcinoma. J Gen Virol 88:1526–1531.

Altindis M, Yilmaz S, Dikengil T, Acemoglu H, Hosoglu S. 2006. Seroprevalence and genotyping of hepatitis B, hepatitis C and HIV among healthy population and Turkish soldiers in Northern Cyprus. World J Gastroenterol 12:6792–6796.

Altuglu I, Soyler I, Ozacar T, Erensoy S. 2007. Distribution of hepatitis C virus genotypes in patients with chronic hepatitis C infection in Western Turkey. Int J Infect Dis 12:239–244.

Ansaldi F, Bruzzone B, Salmaso S, Rota MC, Durando P, Gasparini R, Icardi G. 2005. Different seroprevalence and molecular epidemiology patterns of hepatitis C virus infection in Italy. J Med Virol 76:327–332.

Antaki N, Haddad M, Kebbewar K, Abdelwahab J, Hamed O, Aaraj R, Alhaj N, Haffar S, Assil M, Ftayeh M, Assaad F, Doghman D, Ali T, Nassereldine M, Ali A, Antaki F. 2008. The unexpected discovery of a focus of hepatitis C virus genotype 5 in a Syrian province. Epidemiol Infect 17:1–6.

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

Chamberlain RW, Adams NJ, Taylor LA, Simmonds P, Elliott RM. 1997. The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. Biochem Biophys Res Commun 236:44–49.

Chlabicz S, Flisiak R, Kowalczuk O, Wiercinska-Drapalo A, Pytel-Krolczuk B, Prokopowicz D, Chyczewski L. 2008. High prevalence of genotype 4 among hepatitis C virus-infected intravenous drug users in north-eastern Poland. J Med Virol 80:615–618.

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.

Cochrane A, Searle B, Hardie A, Robertson R, Delahooke T, Cameron S, Tedder RS, Dusheiko GM, De Lamballerie X, Simmonds P. 2002. A genetic analysis of hepatitis C virus transmission between injection drug users. J Infect Dis 186:1212–1221.

Colina R, Casane D, Vasquez S, Garcia-Aguirre L, Chunga A, Romero H, Khan B, Cristina J. 2004. Evidence of intratypic recombination in natural populations of hepatitis C virus. J Gen Virol 85:31–37.

Cristina J, Colina R. 2006. Evidence of structural genomic region recombination in Hepatitis C virus. Virol J 3:53.

de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, Snoeck J, van Rensburg EJ, Wensing AM, van de Vijver DA, Boucher CA, Camacho R, Vandamme AM. 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21:3797–3800.

Esteban JI, Sauleda S, Quer J. 2008. The changing epidemiology of hepatitis C virus infection in Europe. J Hepatol 48:148–162.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Henquell C, Cartau C, Abergel A, Laurichesse H, Regagnon C, De Champs C, Bailly JL, Peigue-Lafeuille H. 2004. High prevalence of hepatitis C virus type 5 in central France evidenced by a prospective study from 1996 to 2002. J Clin Microbiol 42:3030–3035.

Huy TT, Abe K. 2004. Molecular epidemiology of hepatitis B and C virus infections in Asia. Pediatr Int 46:223–230.

Idrees M. 2008. Development of an improved genotyping assay for the detection of hepatitis C virus genotypes and subtypes in Pakistan. J Virol Methods 150:50–56.

Kageyama S, Agdamag DM, Alesna ET, Leano PS, Heredia AM, Abellanosa-Tac-An IP, Jereza LD, Tanimoto T, Yamamura J, Ichimura H. 2006. A natural inter-genotypic (2b/1b) recombinant of hepatitis C virus in the Philippines. J Med Virol 78:1423–1428.

Kalinina O, Norder H, Vetrov T, Zhdanov K, Barzunova M, Plotnikova V, Mukomolov S, Magnius LO. 2001. Shift in predominating subtype of HCV from 1b to 3a in St. Petersburg mediated by increase in injecting drug use. J Med Virol 65:517–524.

Kalinina O, Norder H, Mukomolov S, Magnius LO. 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J Virol 76:4034–4043.

Kamal SM, Nasser IA. 2008. Hepatitis C genotype 4: What we know and what we don't yet know. Hepatology 47:1371–1383.

Katsoulidou A, Sypsa V, Tassopoulos NC, Boletis J, Karafoulidou A, Ketikoglou I, Tsantoulas D, Vafiadi I, Hatzis G, Skoutelis A, Akriviadis E, Vasiliadis T, Kitis G, Magiorkinis G, Hatzakis A. 2006. Molecular epidemiology of hepatitis C virus (HCV) in Greece: Temporal trends in HCV genotype-specific incidence and molecular characterization of genotype 4 isolates. J Viral Hepat 13:19–27.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

Kuiken C, Yusim K, Boykin L, Richardson R. 2005. The Los Alamos hepatitis C sequence database. Bioinformatics 21:379–384.

Kurbanov F, Tanaka Y, Sugauchi F, Kato H, Ruzibakiev R, Zalyalieva M, Yunusova Z, Mizokami M. 2003. Hepatitis C virus molecular epidemiology in Uzbekistan. J Med Virol 69:367–375.

Kurbanov F, Tanaka Y, Avazova D, Khan A, Sugauchi F, Kan N, Kurbanova-Khudayberganova D, Khikmatullaeva A, Musabaev E, Mizokami M. 2007. Detection of hepatitis C virus natural recombinant RF1_2k/1b strain among intravenous drug users in Uzbekistan. Hepatol Res 38:434–457.

Legrand-Abravanel F, Claudinon J, Nicot F, Dubois M, Chapuy-Regaud S, Sandres-Saune K, Pasquier C, Izopet J. 2007. New natural intergenotypic (2/5) recombinant of hepatitis C virus. J Virol 81:4357–4362.

McCaw R, Moaven L, Locarnini SA, Bowden DS. 1997. Hepatitis C virus genotypes in Australia. J Viral Hepat 4:351–357.

Memon MI, Memon MA. 2002. Hepatitis C: An epidemiological review. J Viral Hepat 9:84–100.

Moreau I, Hegarty S, Levis J, Sheehy P, Crosbie O, Kenny-Walsh E, Fanning LJ. 2006. Serendipitous identification of natural inter-genotypic recombinants of hepatitis C in Ireland. Virol J 3:95.

Morice Y, Cantaloube JF, Beaucourt S, Barbotte L, De Gendt S, Goncales FL, Butterworth L, Cooksley G, Gish RG, Beaugrand M, Fay F, Fay O, Gonzalez JE, Martins RM, Dhumeaux D, Vanderborght B, Stuyver L, Sablon E, de Lamballerie X, Pawlotsky JM. 2006. Molecular epidemiology of hepatitis C virus subtype 3a in injecting drug users. J Med Virol 78:1296–1303.

Murphy DG, Willems B, Deschenes M, Hilzenrat N, Mousseau R, Sabbah S. 2007. Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5' untranslated region sequences. J Clin Microbiol 45:1102–1112.

Nakano T, Lu L, Liu P, Pybus OG. 2004. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. J Infect Dis 190:1098–1108.

Naoumov NV. 1999. Hepatitis C virus infection in Eastern Europe. J Hepatol 31(Supp. 1):84–87.

Noppornpanth S, Lien TX, Poovorawan Y, Smits SL, Osterhaus AD, Haagmans BL. 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. J Virol 80:7569–7577.

Pawlotsky JM, Tsakiris L, Roudot-Thoraval F, Pellet C, Stuyver L, Duval J, Dhumeaux D. 1995. Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. J Infect Dis 171:1607–1610.

Pybus OG, Cochrane A, Holmes EC, Simmonds P. 2005. The hepatitis C virus epidemic among injecting drug users. Infect Genet Evol 5:131–139.

Ramia S, Eid-Fares J. 2006. Distribution of hepatitis C virus genotypes in the Middle East. Int J Infect Dis 23:23.

Re V, Contigiani M, Yoshida CF, Lampe E. 2007. Identification of hepatitis C virus subtype 2c by sequencing analysis in patients from Cordoba, Argentina. Mem Inst Oswaldo Cruz 102:995–998.

Richardson D, Fisher M, Sabin CA. 2008. Sexual transmission of hepatitis C in MSM may not be confined to those with HIV infection. J Infect Dis 197:1213–1214; author reply 1214-1215.

Robertson B, Myers G, Howard C, Brettin T, Bukh J, Gaschen B, Gojobori T, Maertens G, Mizokami M, Nainan O, Netesov S, Nishioka K, Shin i T, Simmonds P, Smith D, Stuyver L, Weiner A. 1998. Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: Proposals for standardization. International Committee on Virus Taxonomy. Arch Virol 143:2493–2503.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Samimi-Rad K, Nategh R, Malekzadeh R, Norder H, Magnius L. 2004. Molecular epidemiology of hepatitis C virus in Iran as reflected by phylogenetic analysis of the NS5B region. J Med Virol 74:246–252.

Savvas SP, Koskinas J, Sinani C, Hadziyannis A, Spanou F, Hadziyannis SJ. 2005. Changes in epidemiological patterns of HCV infection and their impact on liver disease over the last 20 years in Greece. J Viral Hepat 12:551–557.

Schroter M, Zollner B, Laufs R, Feucht HH. 2004. Changes in the prevalence of hepatitis C virus genotype among injection drug users: A highly dynamic process. J Infect Dis 190:1199–1200; author reply 1200-1201.

Seeff LB. 2002. Natural history of chronic hepatitis C. Hepatology 36:S35–S46.

Simmonds P. 1999. Viral heterogeneity of the hepatitis C virus. J Hepatol 31(Supp. 1):54–60.

Simmonds P, Holmes EC, Cha TA, Chan SW, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS. 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol 74:2391–2399.

Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, Halfon P, Inchauspe G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin IT, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology 42:962–973.

Tallo T, Norder H, Tefanova V, Krispin T, Schmidt J, Ilmoja M, Orgulas K, Pruunsild K, Priimagi L, Magnius LO. 2007. Genetic characterization of hepatitis C virus strains in Estonia: Fluctuations in the predominating subtype with time. J Med Virol 79:374–382.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analys (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882.

Trepo C, Pradat P. 1999. Hepatitis C virus infection in Western Europe. J Hepatol 31(Supp. 1):80–83.

van Asten L, Verhaest I, Lamzira S, Hernandez-Aguado I, Zangerle R, Boufassa F, Rezza G, Broers B, Robertson JR, Brettle RP, McMenamin J, Prins M, Cochrane A, Simmonds P, Coutinho RA, Bruisten S. 2004. Spread of hepatitis C virus among European injection drug users infected with HIV: A phylogenetic analysis. J Infect Dis 189:292–302.

van de Laar TJ, van der Bij AK, Prins M, Bruisten SM, Brinkman K, Ruys TA, van der Meer JT, de Vries HJ, Mulder JW, van Agtmael M, Jurriaans S, Wolthers KC, Coutinho RA. 2007. Increase in HCV incidence among men who have sex with men in Amsterdam most likely caused by sexual transmission. J Infect Dis 196:230–238.

Verbeeck J, Maes P, Lemey P, Pybus OG, Wollants E, Song E, Nevens F, Fevery J, Delport W, Van der Merwe S, Van Ranst M. 2006. Investigating the origin and spread of hepatitis C virus genotype 5a. J Virol 80:4220–4226.

Watson JP, Al-Mardini H, Awadh S, Ukabam S, Record CO. 1999. Hepatitis C virus genotypes in a cohort of Middle Eastern patients. Ann Saudi Med 19:410–412.

WHO. 1999. Global surveillance and control of hepatitis C. Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium. J Viral Hepat 6:35–47.

Zein NN. 2000. Clinical significance of hepatitis C virus genotypes. Clin Microbiol Rev 13:223–235.

Zeuzem S. 2004. Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: Who responds less well? Ann Intern Med 140:370–381.