

**Genetic Diversity and Molecular evolution of
Hepatitis C Virus**

Suwanna Noppornpanth

The research described in this thesis was conducted at the Department of Virology of the Erasmus Medical Center in Rotterdam, The Netherlands. Parts of the studies were funded by the Commission of the European Community, HECSA Project (grant number ICA4-CT-1999-10009).

Printing of this thesis was financially supported by:
Novartis Vaccine & Diagnostic SRL, Siena, Italy.

Design of the cover : Salin & Suwanna

ISBN : 978-90-9023431-1

Printed by: PrintPartners Ipskamp, Rotterdam, The Netherlands

© S. Noppornpanth, The Netherlands, 2008. All rights reserved. No part of this thesis may be reproduced or transmitted, in any form or by any means, without prior written permission of the author.

**Genetic Diversity and Molecular Evolution of
Hepatitis C Virus**

**Genetische diversiteit en moleculaire evolutie van
hepatitis C virus**

Thesis

**to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the rector magnificus**

Prof. dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board

**The public defense shall be held on
Wednesday 24 September 2008 at 9:45 hrs**

by

**Suwanna Noppornpanth
Born in Bangkok, Thailand**



Doctoral Committee

Promoter :

Prof.dr. A.D.M.E Osterhaus

Other members :

Prof.dr. C.A.B. Boucher

Prof.dr. S. Abrignani

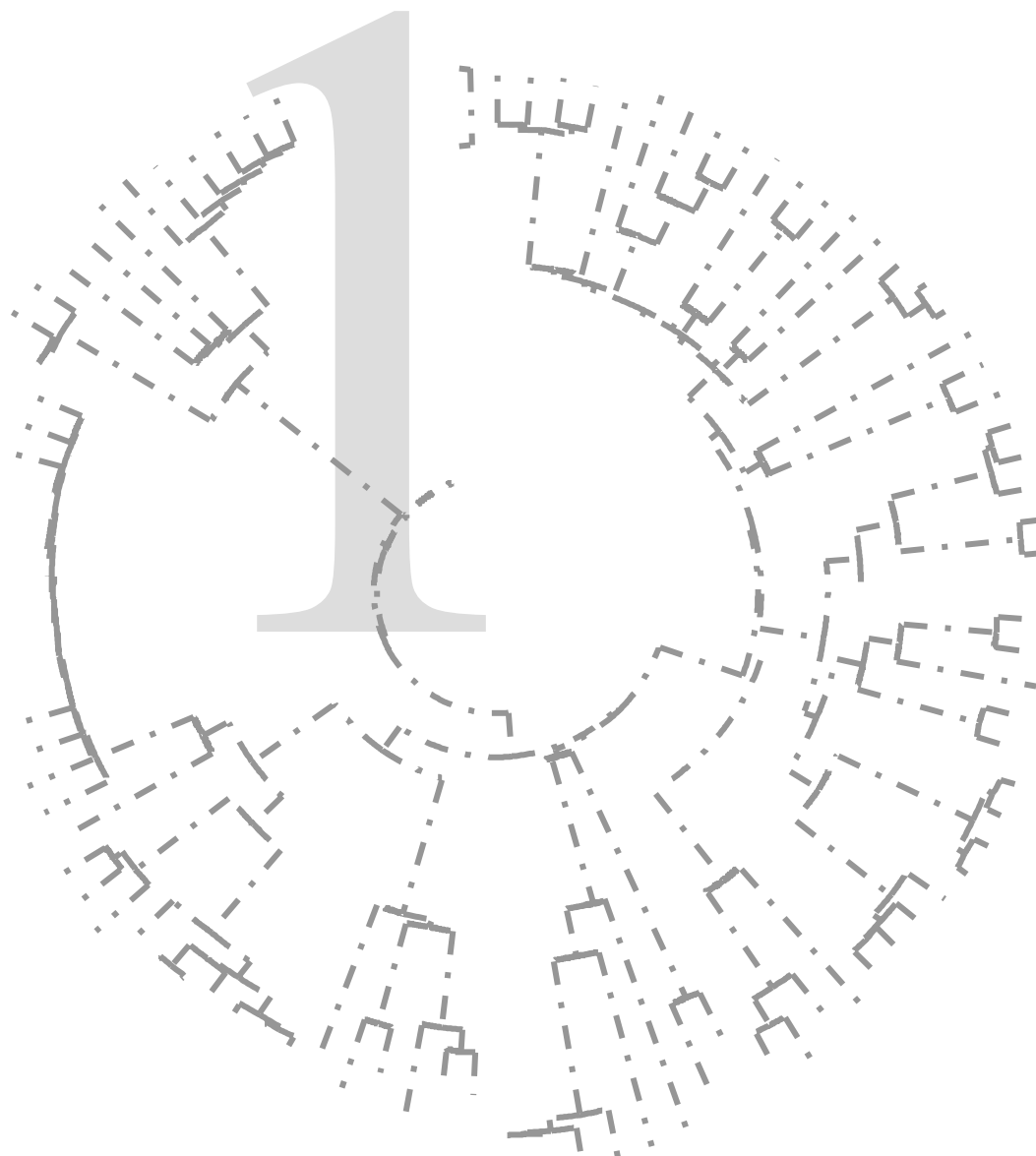
Prof.dr. R.A.M. Fouchier

Copromoter :

Dr. B.L. Haagmans

Table of Contents

Chapter 1	General Introduction	10
Chapter 2	Genotyping hepatitis C viruses from Southeast Asia by a novel line probe assay that simultaneously detects core and 5' untranslated regions. Journal Clinical Microbiology (2006)	36
Chapter 3	Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. Journal of Virology (2006)	54
Chapter 4	Characterization of hepatitis C virus deletion mutants circulating in chronically infected patients. Journal Virology (2007)	76
Chapter 5	Complete genome analysis of hepatitis C virus subtypes 6t and 6u. Journal General Virology (2008)	96
Chapter 6	Molecular evolution of Hepatitis C Viruses : dating the origin of viruses using phylogenetic and relaxed molecular clock analysis. Manuscript in preparation	118
Chapter 7	Detection of hepatitis C virus specific T cells using Epstein Barr virus based plasmids. Manuscript in preparation	140
Chapter 8	Summarizing discussion	158
Nederlandse samenvatting		182
Acknowledgements		190
Curriculum Vitae		193
Publications		194



CHAPTER 1

GENERAL INTRODUCTION

Hepatitis C virus (HCV), an enveloped positive stranded RNA virus, is the causative agent of non-A, non-B (NANB) hepatitis (27). The virus was identified and characterized by molecular cloning techniques using serum from a NANB hepatitis virus infected chimpanzee (15) and based on the similarity of the genome organization and hydropathy profiles of several precursor proteins classified as a member of the *Flaviviridae* family. However, the low sequence homology compared to other flaviviruses eventually lead to its classification into a new genus *hepacivirus*, distinct from the other flavivirus members (33, 74).

Initial studies demonstrated that blood transfusions were the main transmission route that caused the HCV epidemic, especially prior to the period of HCV-contaminated blood screening in 1992 (50). HCV high-risk groups include recipients of multiple blood transfusion, hemophiliacs, hemodialysis patients and intravenous drug users. Modes of transmission include contact with contaminated-blood products, sharing needles, mother-to-child and sexual transmissions (69). Infection of the human liver is often clinically benign, with mild symptoms in the acute phase and fulminant hepatitis is very rare. The disease may even go unnoticed in cases of acute resolving HCV. In most cases (>70%) HCV infection leads to chronic persistent or active infection, often accompanied by complications of liver cirrhosis, or type II cryoglobulinaemia. The mechanisms of HCV persistence are not fully understood but the inability of the host innate and specific immune system to clear the virus is remarkable for a positive stranded RNA virus. HCV specific CD8⁺ T-cell responses are considered to play an important role in the containment and clearance of HCV infection but few of these cells are present in the liver and are unable to eradicate the infection. In addition, antibodies to HCV do not seem to play a major role in clearance of infection and re-infections are often observed with a different HCV strain, suggesting that immune responses to the initial strain did not protect against an infection with another strain of HCV (43).

HCV infection has reached epidemic proportions. Globally, an estimated 170 million individuals are chronically infected and more than one million new infection cases are reported annually (19, 106). In the United States alone, nearly four million persons are infected and 30,000 acute new infections are estimated to occur each year (3). In Europe and Japan, the disease is already more important numerically than is either hepatitis B virus (HBV) or human immunodeficiency virus (HIV) infection and due to

the availability of the HBV vaccine the impact of hepatitis C infections will increase further. HCV infection causes a substantial portion of chronic liver disease mortality due to the induction of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (HCC) (52). About 4 to 20% of patients with chronic hepatitis C will develop liver cirrhosis within 20 years and HCC may develop after about 20 to 35 years. There is a significant rise in the incidence of HCC in many developed countries including Japan, Spain, France, and Italy, where the proportion of HCC attributable to HCV ranges from 50% to 70% (56). In Japan, HCV-related HCC incidence has more than tripled over the past four decades and accounts in the 60–70 year age group for as much as 90% (38). Apart from HCC, co-infections with other viruses, especially HIV-1 and other hepatitis viruses, have gained more attention. These are of clinical importance, since the course of HCV infection is accelerated by co-infection with HIV-1 (75, 89), hepatitis A virus (HAV) or HBV (101, 108).

HCV genome organization and viral replication

The HCV genome consists of an RNA molecule of approximately 9,600 nucleotides structured in a coding region that contains one large open reading frame, flanked by non-translated regions (UTR) at the 5' and 3' end. The polyprotein is cleaved into the structural (core, E1 and E2) and nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A and NS5B) with one additional small protein at the junction between the structural and non-structural elements (p7 protein) (Figure 1, (47)). The 5'untranslated region (5'UTR) consists of 341 nucleotides with highly conserved sequences among different HCV-isolates and contains an internal ribosome entry site (IRES). The IRES binds directly to the 40S ribosomal subunit controlling the viral protein translation, and is also important in viral pathogenesis (67, 99). The 3'untranslated region (3'UTR) contains the typical poly A or poly U, followed by a unique 98 nucleotide sequence to protect genomic RNA from degradation and aid in translation. This latter tail sequence was highly conserved among virus strains and can fold into a conserved stem-loop structure (39, 94). It suggests that the 3'UTR is a structure common to all HCV genomes and plays an important role in the initiation of viral replication.

The polyprotein is translated as one open reading frame and subsequently post-translationally processed by both viral and cellular proteinases. Processing at the core/E1,

E1/E2, E2/p7, and p7/NS2 sites by host cell signal peptidase produces all structural proteins and p7. Two viral proteases are responsible for the maturation of the nonstructural proteins. NS2 contains an autoprotease, which cleaves the junction between NS2 and NS3 while NS3 protease cleaves the remaining non-structural proteins with NS4A as a cofactor for this activity (Figure 1).

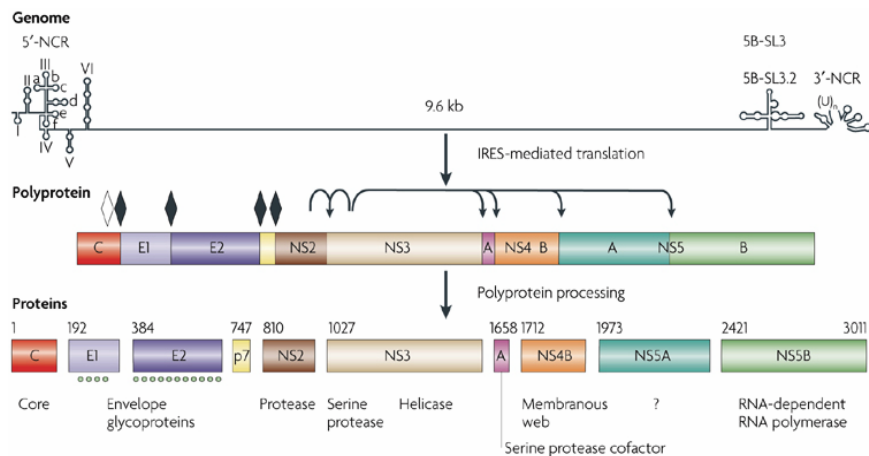


Figure 1 HCV genome organization. The 9.6-kb positive-strand RNA genome is schematically depicted at the top. Simplified RNA secondary structures in the 5'- and 3'-non-coding regions (NCRs) and the core gene, as well as the NS5B stem-loop 3 *cis*-acting replication element (5B-SL3) are shown. Internal ribosome entry site (IRES)-mediated translation yields a polyprotein precursor that is processed into the mature structural and non-structural proteins. Amino-acid numbers are shown above each protein (HCV H strain; genotype 1a; GenBank accession number [AF009606](#)). Solid diamonds denote cleavage sites of the HCV polyprotein precursor by the endoplasmic reticulum signal peptidase. The open diamond indicates further C-terminal processing of the core protein by signal peptide peptidase. Arrows indicate cleavages by the HCV NS2–3 and NS3–4A proteases. Dots in E1 and E2 indicate the glycosylation of the envelope proteins (4 and 11 *N*-linked glycans, respectively, in the HCV H strain) (57).

The HCV CORE protein is a highly basic, RNA-binding protein that presumably forms the viral nucleocapsid. It has been reported that CORE protein interacts with numerous cellular proteins and affects host cell functions such as gene transcription, lipid metabolism, apoptosis and various signaling pathways (96). In addition, CORE has been associated with the induction of steatosis and HCC development (45). Several groups reported the existence of a protein encoded by an alternative reading frame within the core region (8). It comprises up to 160 aminoacids and is designated as ARFP (alternative reading frame protein) or F (frameshift) protein. The ARFP/F protein is dispensable for HCV RNA replication and whether it is expressed during natural HCV infection remains to be clarified.

The envelope glycoproteins, E1 and E2, are type I transmembrane proteins with C-terminal hydrophobic anchors, which mediate entry of the virus into the host cell. These E1 and E2 proteins form non-covalent heterodimers that represent the structure on the viral envelope confirmed by earlier electron microscopy observations and structural studies of the recombinant HCV particles, even though the particle assembly and release processes are poorly understood (102).

p7 is a 63-amino acid polypeptide located at the junction between the structural and nonstructural region. It is composed of two transmembrane domains and forms hexamers with ion channel activity (65). It is believed that p7 could be important for viral assembly because the corresponding protein of the related bovine viral diarrhea virus (BVDV) is essential for the production of infectious progeny virus but not for RNA replication (32).

Although NS2-3 protease activity is required for the replication *in vivo*, it is dispensable for replication of subgenomic replicons *in vitro* (68). Interestingly, the NS2-3 region is the fusion site for the construction of efficient HCV chimeras (4) and natural occurring recombinant viruses (36).

NS3 is a multifunctional protein as it contains a serine protease located in the N-terminal one-third that is responsible for the downstream cleavage in the nonstructural region and a NTPase/RNA helicase domain in the C-terminal two-thirds. Together with NS4A, a 54-amino acid polypeptide that is required as a cofactor, the NS3 serine protease enzyme is localized in intracellular membranes. The enzymatic activity of the NS3 NTPase/helicase is essential for RNA replication. Putative functions during replication could be to unravel replicative double strand RNA intermediates, to eliminate RNA secondary structures or to separate the genome from nucleic acid binding proteins. Recently, the NS3 serine protease turned out to influence the innate cellular host defense by inhibiting RIG-I and TLR3 signaling (29, 55). Therefore, the NS3 protease is particularly attractive target for the development of new antivirals (21, 76).

NS4B is a 27-kDa integral membrane protein that localizes to an ER-derived membranous compartment (35). Interestingly, the expression of NS4B induces a specific membrane alteration, designated as membranous web, and serves as a scaffold for the formation of the viral replication complex (31).

NS5A is a phosphorylated zinc metalloprotein, that has attracted considerable interest because of its potential role in modulating the IFN response (93), even though these findings are still controversial. A striking observation was the concentration of cell culture adaptive replicon mutations within the central part of NS5A (7). These observations support the concept that NS5A plays an important role in the regulation of viral replication (48). The NS5B RNA dependent RNA polymerase (RdRp) is the key enzyme for HCV replication that promotes synthesis of new RNA genomes. Replication proceeds *via* synthesis of a complementary minus-strand RNA using the genome as a template and the subsequent synthesis of genomic plus strand RNA from this minus-strand RNA intermediate. Due to its influence on HCV replication, NS5B has emerged as a major target for antivirals (21).

Circulating HCV particles can be associated with low- and very-low-density lipoproteins (LP) and virus binding to the cell surface and entry may involve the low density lipoprotein receptor (LDLR). In addition several other membrane proteins, such as glycosaminoglycan (GAG), scavenger receptor class B type I (SR-B1), the tetraspanin protein CD81 and Claudin-1 (CLDN1), are thought to be involved in entry of virus. HCV entry likely involves transit through an endosomal, low pH compartment and fusion with the endosomal membrane (57). Subsequently, the positive-strand RNA genome serves as a template to make a negative-strand intermediate, which then serves as a template to produce multiple nascent genomes (Figure. 2). As in all positive-strand RNA viruses investigated so far (1), HCV forms a membrane associated replication complex, composed of viral proteins, replicating RNA, altered cellular membranes and additional host cell factors (24, 31). This specific membrane alteration, called the membranous web, was identified as the site of RNA replication in Huh-7 cells harboring subgenomic HCV replicons (31). In addition to coordinated protein-protein and protein-membrane interaction, essential *cis*-acting replication elements (CRE) of the RNA genome were recently discovered (107). Little is known about the late steps of the viral lifecycle, as these have only recently become available for systematic study. Virions presumably form by budding into the ER, or an ER-derived compartment, and exit the cell through the secretory pathway. A possible link between lipoprotein metabolism and viral assembly and release has been proposed (34).

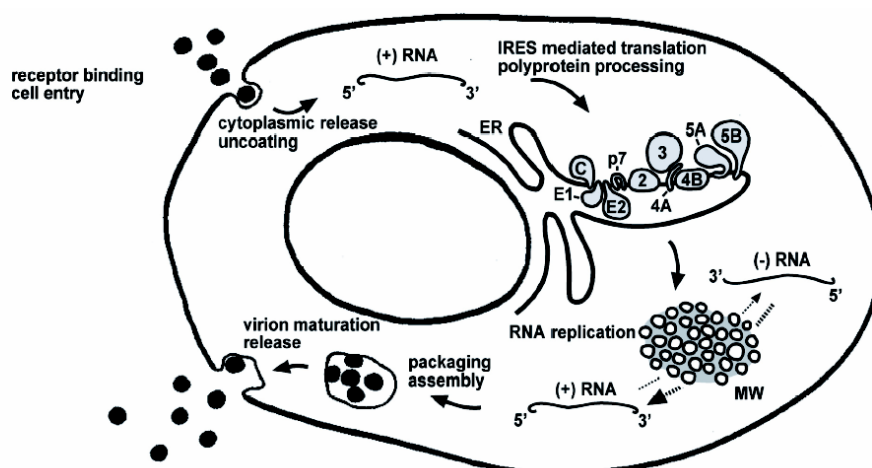


Figure 2. Life cycle of HCV. The steps of the viral life cycle are depicted schematically. The topology of HCV structural and nonstructural proteins at the endoplasmic reticulum (ER) membrane is shown. HCV RNA replication occurs in a specific membrane alteration, the membranous web (MW). IRES-mediated translation and polyprotein processing as well as membranous web formation and RNA replication, illustrated here as separate steps for simplicity, may occur in a tightly coupled manner (9).

Although exciting progress has been made to determine the virion structure of the related alphaviruses and flaviviruses, HCV has not been definitively visualized and its structure remains to be elucidated. Based on filtration and electron microscopic studies, the HCV particle is a spherical lipid envelop with an approximate diameter of approximately 70 nm (102). The nucleocapsid of the particles, which were 50 nm in diameter, appeared to be icosahedral in structure and surrounded by an enveloped covered with surface projection (46).

For many years, the extremely restricted host range and the inefficiency of *in vitro* models based on the incubation of cells in culture with HCV particles, hampered HCV research. A breakthrough came in 1999 when Lohmann *et al.* established the first functional subgenomic replicon system for HCV (49). In this system, the structural region was replaced by the sequence encoding the neomycin phosphotransferase gene resulting in G418 resistance and expression of the nonstructural proteins is directed by a heterologous encephalomyocarditis virus (EMCV) internal ribosomal entry site (IRES) downstream of the neomycin phosphotransferase gene. After transfection into the human HCC cell line Huh-7, G418 resistant colonies contain self-replicating HCV RNA were selected. The replicon system made it possible to study efficient and genuine HCV RNA replication *in vitro* and structural aspects of the replication complex, virus-host cell interactions, and the efficacy of antiviral agents. A few years later, Wakita and

colleagues, generated a genotype 2a replicon (JFH-1), isolated from the serum of a patient with fulminant hepatitis C (102). This system turned out to replicate very efficiently in different cell types. Furthermore, the full-length JFH-1 sequence produced infectious viral particles that could be passaged in cell culture (102). Newly constructed chimeras with the structural region of the J6 genotype 2a clone improved this system significantly (47). The infectious HCV cell culture system represents the last major milestone in the field and renders the complete viral life cycle accessible to detailed analysis *in vitro* and may allow the identification of novel targets for antiviral therapy.

Genomic heterogeneity and classification systems of HCV

The HCV genome shows remarkable sequence variation generated by the lack of proofreading activity by the RNA polymerase and a high *in vivo* productive rate of an estimated 10^{12} virions per day (61). To encompass this diversity, HCV has been classified in genotypes and subtypes (10, 97). Initially, genotypes were considered to differ more than 20% at the nucleotide level and more than 15% at the amino acid level (37). However, phylogenetic analyses indicated that genotype 10 is closely related to genotype 3 and genotype 7, 8, 9 and 11 to genotype 6 (82). Therefore, a new consensus nomenclature system was proposed to be used for HCV classification (81). According to this system, HCV is classified into genotypes on the basis of < 70% similarity of nucleotide sequence and phylogenetic relationship. The more closely related HCV strains (75-80% sequence similarity) within some genotypes are designated subtypes. Genotypes are numbered in order of discovery and subtypes are assigned lowercase letters (Table 1). Nowadays, HCV is classified into 6 major genotypes and more than 90 subtypes. Sequence variation between the HCV genomes was on average 0.9% (53), indicative of the presence of HCV quasispecies in hepatitis C patients. Furthermore, sequence heterogeneity was distributed throughout the genome except in the 5'UTR, while diversity was higher (1.6%) in the envelope genes than the others (95). A comparison of the complete HCV genome reveals that non-synonymous substitutions are less frequent compared to synonymous ones, except in hypervariable region 1 (HVR1). Moreover, translations are more common than transversions, particularly at the third position of the codon (86). Currently, intensive sequencing of HCV genomes is being conducted and nearly 50,000 HCV sequences, including 507 full-length genomes, have been deposited

in generic data banks so far (41). To perform sequence classification, annotation, and analysis of such a large collection, several sequence databases are dedicated specifically to HCV, such as Los Alamos HCV database (<http://hcv.lanl.gov/content/hcv-index>), the hepatitis C virus database (<http://s2as02.genes.nig.ac.jp/>) and HCVDB (<http://euhcvdb.ibcp.fr/euHCVdb/jsp/index.jsp>).

Table 1 Genotypes/subtypes that are presently defined in the database

Genotype	Assigned subtypes
1	a, b, c, d, e, f, g, h, i, j, k, l, m
2	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r
3	a, b, c, d, e, f, g, h, i, k, l
4	a, b, c, d, e, f, g, h, k, l, m, n, o, p, q, r, s, t
5	a
6	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t

HCV genotyping

HCV genotype identification is clinically important for prediction of response to, and in determining the duration of antiviral therapy (109). Nowadays, most treatment protocols require preceding genotype information for HCV infected patients. This is illustrated by the fact that genotype 1 and 4 are more resistant to treatment with pegylated alpha interferon and ribavirin than genotype 2 and 3 (54, 110). Moreover, it has been suggested that patients with chronic HCV genotype 1b infection show more severe liver disease than patients infected with other genotypes (78).

HCV genotyping based on the detection of genotype-specific HCV antibodies is relatively easy to perform providing a tool for large epidemiological studies. One of the serologic genotyping assays contains five different specific peptide sequences of the NS4 region and two specific peptide of the core region taken from the HCV genomes of genotypes 1, 2, and 3 (22). The second serologic genotyping assay is based on the detection of genotype-specific antibodies directed to epitopes encoded by the NS4 region of the genomes for genotypes 1 through 6. Results from both assays showed a concordance rate of more than 96% for genotypes 1, 2, and 3 (30), but low sensitivity of

genotyping has been observed in other studies using sera from HCV genotype 6 or HIV coinfecting patients (5, 88).

Molecular genotyping techniques to detect HCV RNA rely on the amplification of short HCV RNA regions from clinical specimens, followed by a HCV genotype-specific assay. The gold standard and most definitive method for HCV genotyping is sequencing of a specific PCR-amplified portion of the HCV genome obtained from the patient, followed by phylogenetic analysis. Determination of HCV genotype has been focused on sequence analysis of the HCV complete genome, NS5b, Core, E1, and 5'UTRs (83). Other methods that have been reported lately depend on the short amplification of HCV RNA, followed by re-amplification with type-specific primers (64), hybridization with type-specific probes (91) or by digestion of PCR products with restriction endonucleases that recognize genotype-specific cleavage site (11). The VERSANT[®] HCV genotype assay, (INNO-LiPA) is based on the amplification of the 5'UTR with biotinylated primers, after which the PCR product is hybridized to a membrane impregnated with genotype-specific probes and detected with streptavidin linked to a colorimetric detector (90). However, it has been noted that this test directed towards the 5'UTR falsely identifies some genotype 6 variants (6c-6l) from South East Asia (SEA) as genotype 1 (13, 91) and it is unable to distinguish subtypes 1a from 1b in 5-10% of the cases (12, 84). In general, all molecular HCV genotyping methods have a high sensitivity and specificity. Sequencing of amplified DNA, however, does not usually identify mixed infections with two different HCV genotypes. Even though all these methods are suitable to identify precisely the major HCV genotypes, only the direct nucleotide sequencing followed by phylogenetic analysis is efficient in discriminating among subtypes. However, these techniques are costly, time-consuming and require specialized facilities to ensure accurate results.

Epidemiology and geographic distribution of HCV

HCV shows a large degree of geographic variability in its distribution. Countries with the highest reported prevalence rates are located in Africa and Asia; areas with lower prevalence include the developed nations in North America, Northern and Western Europe, and Australia (Fig.3) (77). The seroprevalence rate is about 1% in Western Europe and North America, 3-4% in some Mediterranean and Asian countries and up to

10-20% in parts of Central Africa and Egypt (103). Populous nations in the developed countries with relatively low rates of HCV seroprevalence include Germany (0.6%), Canada (0.8%), France (1.1%), and Australia (1.1%). Low, but slightly higher seroprevalence rates have been reported in the USA (1.8%), Japan (1.5–2.3%), and Italy (2.2%) (for a review see (77)).

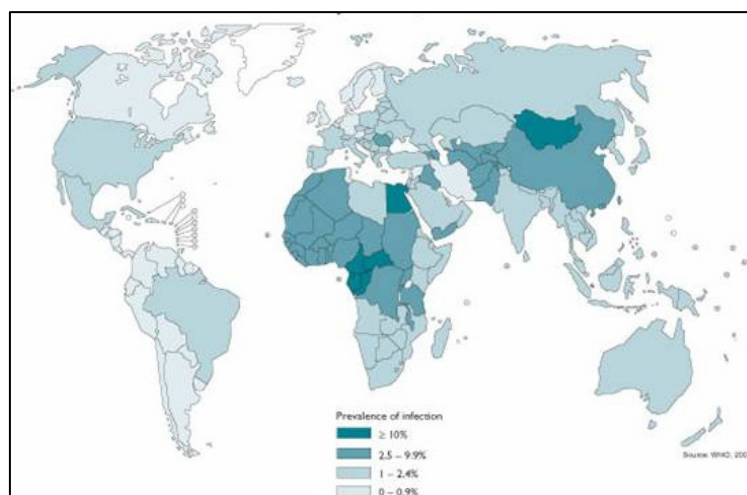


Figure 3 Prevalence of HCV in the world (Source, WHO 2007)

(http://www.nathnac.org/pro/factsheets/images/hep_c_epi.jpg)

There is a wide range of prevalence estimates from developing countries, and generally less data are available to validate assumptions about the burden of disease compared with the developed world. China, whose citizens account for one fifth of the world's population, has a reported seroprevalence of 2.5-4.9% (105). In other countries like India, Indonesia and Pakistan, this may range from 0.9 % to 6.5% (16, 58, 92). Egypt, with an estimated population of 73 million, has the highest reported seroprevalence rate at 22% due to the use of parenteral antischistosomal therapy contaminated with HCV (28).

HCV genotypes have a geographic distribution with genotype 1 being more common in the Americas and Europe, followed by genotypes 2, 3 and 4. Subtype 1b accounts for most infections in China and Japan, genotype 4 is more prevalent in the Middle East and North Africa, genotype 5a is seen almost exclusively in South Africa, and genotype 6 is common in Hong Kong and South-East Asia (23, 79). Variants of

genotype 6, originally termed genotypes 7, 8 and 9, have been reported primarily in South-East Asia and genotypes 10 and 11 in Indonesia (97, 98).

Through a phylogenetic analysis of nucleotide sequences of the HCV genome, such as the NS5B region, it is possible to demonstrate clustering of HCV isolates that are more or less closely related, even within a single subtype. Using these techniques spread of HCV genotype 1b through infected blood products and subtype 1a and 3a in intravenous drug users, was demonstrated (6, 66). Together with the coalescent theory of population genetics, phylogenetic analysis has been used to estimate the historical age and distribution rate of different HCV genotypes (59, 70). As a result, these models indicate that genotype 1 originated within the last 100 years, whereas types 4 and 6 are several times older. Subtype diversity within genotypes 1, 2 and 4 from Western Africa and genotypes 3 and 6 in South-East Asia implicates long-term presence of HCV in these populations (59, 70). Based on the genotype distributions, it was suggested that HCV has been endemic in sub-Saharan Africa and South-East Asia for a considerable time, and that the occurrence of infection in Western and other non-tropical countries represents a relatively recent emergence in new risk groups (60, 80). In the 20th century, widespread use of blood transfusion, unsterilized needles for injections and vaccinations caused HCV to spread. These new transmission routes account for the epidemic spread of HCV over the past 50 years in Europe, Egypt and elsewhere (17, 60, 70, 71).

HCV molecular evolution

Dating the origin of HCV and estimate when HCV was introduced into human population remains difficult due to the inability to identify HCV or HCV-like variants in ape or monkey species (51). In theory, it may be possible to calculate the divergence times of the main clades and splitting of subtypes by using the constant nucleotide substitution rate over time. By comparing HCV sequences from sequential samples from chronically infected individuals or from those infected by a common source, rates of sequence change were measured to be 1.44×10^{-3} nucleotide changes per site per year over the whole genome (63). Smith et al. calculated an evolutionary rate of 7.4×10^{-4} nucleotide substitutions per site per year for the E1 gene and 4.1×10^{-4} for the NS5B gene and suggested that the subtypes diverged around 300 years ago, and the divergence of the different genotypes should have occurred more than 500-2000 years ago (85).

However, it is known that HCV is a fast evolving virus and does not contain a constant rate of nucleotide substitution sites along the genome. Therefore, the evolutionary dynamics of highly diverse genetic groups, such as HCV genotype 6, still remain elusive.

The forces that drive the early divergence of the six major genotypes of HCV and the underlying basis of the current geographical distribution of genotypes also remain unclear. Hypothetically, HCV evolution may be driven by sequence variation caused by the lack of proofreading activity of the viral RNA polymerase RdRP creating a quasispecies and further amplified by immune selection. Because genome changes occur frequently at synonymous sites in the coding regions of the genome (86), it is unlikely that viruses would accumulate phenotypically neutral changes. The expectation from this neutral theory would be that such diversification should occur at a constant rate over time, allowing to estimate spread of HCV in specific transmission networks, such as IDUs (17, 70, 71). The prediction of the time of introduction of specific genotypes, such as 1a, 1b, 3a and 4a, into new risk groups for infection in Western countries was estimated in the same way (87). The phylogenetic tree structure of genotype 4a in Egypt demonstrated that it is compatible with the time of introduction of HCV into that population through potential treatment for schistosomiasis (*Bilharzia*) in the 1950s and 1960s (28, 71, 72). The increasing sequence diversity within genotypes 3a, 1a, 1b, 2a and 2b, respectively, suggests time points of introducing these viruses at earlier times in the 20th century, associated with other risk factors for infection, such as injecting drug use, blood transfusion and large-scale immunizations.

Recombination was reported in many families of RNA viruses, including other members of the family Flaviviridae (44, 100). It's role in HCV evolution became evident only a decade after HCV was discovered, through the identification of a naturally occurring inter-genotypic recombinant form of HCV (1R_1b/2k) in St. Petersburg (36). More recently, an intragenotypic recombinant HCV (1a/1b) has identified in Peru (18). The true frequency of HCV recombination may be underestimated, especially in areas of endemic with a high prevalence of different HCV genotypes, due to currently used methods that are not suited for identification of recombinant strains (18, 79).

Viral escape from cytotoxic T-cells lymphocytes (CTLs) and antibodies has been identified as a major driving force for the evolution of human and simian immunodeficiency virus but also HCV (2, 62). In HCV infection, two recent studies

support a role for CD8 T-cell responses and reversions in driving HCV evolution (20, 73). Both studies identified dominant viral escape mutants in HCV patients with persistent viremia. Amino acid substitutions occurred 13-fold more often within than outside T cell epitopes, and every chronic HCV patient carried amino acid substitution in at least one epitope while there was no substitution in T cell epitopes for a year after infection in a person who cleared viremia. Similarly, studies in HCV infected chimpanzees demonstrated that over a period of several years a new strain of the virus emerged with a mutation in the CTL epitope that was no longer recognized by the CTLs isolated earlier (40). Apart from escaping CTL responses, several studies have shown B-cell epitope shifts, apparently the result of the presence of anti-hypervariable region (HVR) antibodies, which support the hypothesis that antibody neutralization allows selection of variants (26, 42, 104). Neutralization of HCV in chimpanzees with homologous plasma obtained from the same chronically infected patient showed that neutralization was achieved with plasma obtained 2 years after the initial exposure but not with plasma obtained 11 years later (25). Analysis of viral isolates of the same patient showed significant genetic divergence of HCV over time.

Aim and the outline of the thesis

The goal of this thesis was to characterize and study genetic diversity and molecular evolution of HCV in South East Asia (SEA). The enormous HCV sequence diversity represents a significant hurdle to develop effective vaccines and novel therapeutic interventions. The current standard treatment of hepatitis C is combination of pegylated interferon- α and ribavirin that can induce a sustained response in 42–82% of patients, although therapy must be discontinued in some patients due to side effects associated with treatment. Clinical studies have shown that HCV genotype is an independent prognostic factor in predicting response to antiviral therapy (78). In addition, vaccine development has been slowed by the extreme antigenic variability of the virus, although recombinant E1 and E2 protein vaccines can elicit protective immunity when tested in chimpanzees (14).

To study prevalence and incidence of different HCV genotype infections in Thai and Vietnamese urban settings, serum samples were collected from selected populations, with different risks of being infected within the setting of the EU funded project HECSA

(Hepatitis C vaccine efficacy against South East Asian genotypes). Careful monitoring of the epidemic situation in a given population is one of the important issues in vaccine development and in preparing for clinical trials. In both countries, samples were collected retrospectively and prospectively from regular blood donors with a track record of regular blood donation and hepatitis C patients. All samples were screened by commercial ELISA for antibodies against HCV and used for molecular characterization of the virus present. The Thai laboratories, in close collaboration with the European laboratories, subjected positive samples to detailed genotype analysis using standard RT-PCR sequencing protocols. Special attention was paid to the identification of genotype 6 positive samples and the occurrence of subtypes within this genotype. HCV genotype 6 circulate specifically in Asia and SEA, but rare in other part of the world. Information about epidemiology, the effect of diversity on clinical aspects and the molecular evolution of HCV genotype 6 are limited. In this thesis, several studies were designed to increase our knowledge regarding the HCV molecular evolution in SEA by identification of blood donors' HCV-positive sera collected from Bangkok, Thailand and Ho Chi Minh City, Vietnam.

Because molecular diagnosis based on the conserved region (5'UTR) could not distinguish HCV genotype 1 and 6, a new developed diagnostic test to separate genotype 6 from 1 based on 5'UTR-Core genes was examined (Chapter 2). In addition, through analysis of two distinct regions in the HCV genome, a natural recombination of different HCV genomes (genotype 2i/6p) was found in one sample collected from Vietnam (chapter 3). Additionally, HCV defective interfering (DI) viruses with a large fragment deletion of E1-E1-NS2 regions were found to be present in the plasma of chronic infected patients (Chapter 4). Both mechanisms may further support the idea that variation of HCV is caused through different molecular mechanisms. Next the molecular evolution of HCV genotypes 1-6 was studied using whole genomes available from the HCV database and those characterized in this study (Chapter 5 and 6). Because host CTL responses play an important role in the molecular evolution of HCV, tools were developed to detect CTL activity in chronically infected patients using a plasmid system enabling the expression of diverse HCV genome fragments (Chapter 7). The findings presented in this thesis are evaluated in a summarizing discussion (Chapter 8).

REFERENCES

1. **Ahlquist, P., A. O. Noueir, W. M. Lee, D. B. Kushner, and B. T. Dye.** 2003. Host factors in positive-strand RNA virus genome replication. *J Virol* **77**:8181-6.
2. **Allen, T. M., M. Altfeld, S. C. Geer, E. T. Kalife, C. Moore, M. O'Sullivan K, I. Desouza, M. E. Feeney, R. L. Eldridge, E. L. Maier, D. E. Kaufmann, M. P. Lahaie, L. Reyor, G. Tanzi, M. N. Johnston, C. Brander, R. Draenert, J. K. Rockstroh, H. Jessen, E. S. Rosenberg, S. A. Mallal, and B. D. Walker.** 2005. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J Virol* **79**:13239-49.
3. **Armstrong, G. L., A. Wasley, E. P. Simard, G. M. McQuillan, W. L. Kuhnert, and M. J. Alter.** 2006. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* **144**:705-14.
4. **Bartenschlager, R., and T. Pietschmann.** 2005. Efficient hepatitis C virus cell culture system: what a difference the host cell makes. *Proc Natl Acad Sci U S A* **102**:9739-40.
5. **Beld, M., M. Penning, M. van Putten, A. van den Hoek, V. Lukashov, M. McMorro, and J. Goudsmit.** 1998. Hepatitis C virus serotype-specific core and NS4 antibodies in injecting drug users participating in the Amsterdam cohort studies. *J Clin Microbiol* **36**:3002-6.
6. **Berg, T., U. Hopf, K. Stark, R. Baumgarten, H. Lobeck, and E. Schreier.** 1997. Distribution of hepatitis C virus genotypes in German patients with chronic hepatitis C: correlation with clinical and virological parameters. *J Hepatol* **26**:484-91.
7. **Blight, K. J., A. A. Kolykhalov, and C. M. Rice.** 2000. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**:1972-4.
8. **Branch, A. D., D. D. Stump, J. A. Gutierrez, F. Eng, and J. L. Walewski.** 2005. The hepatitis C virus alternate reading frame (ARF) and its family of novel products: the alternate reading frame protein/F-protein, the double-frameshift protein, and others. *Semin Liver Dis* **25**:105-17.
9. **Brass, V., D. Moradpour, and H. E. Blum.** 2006. Molecular virology of hepatitis C virus (HCV): 2006 update. *Int J Med Sci* **3**:29-34.
10. **Bukh, J., R. H. Miller, and R. H. Purcell.** 1995. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Semin Liver Dis* **15**:41-63.
11. **Buoro, S., S. Pizzighella, R. Boschetto, L. Pellizzari, M. Cusan, R. Bonaguro, C. Mengoli, C. Caudai, M. Padula, P. Egisto Valensin, and G. Palu.** 1999. Typing of hepatitis C virus by a new method based on restriction fragment length polymorphism. *Intervirology* **42**:1-8.
12. **Chen, Z., and K. E. Weck.** 2002. Hepatitis C virus genotyping: interrogation of the 5' untranslated region cannot accurately distinguish genotypes 1a and 1b. *J Clin Microbiol* **40**:3127-34.
13. **Chinchai, T., J. Labout, S. Noppornpanth, A. Theamboonlers, B. L. Haagmans, A. D. Osterhaus, and Y. Poovorawan.** 2003. Comparative study of different methods to genotype hepatitis C virus type 6 variants. *J Virol Methods* **109**:195-201.

14. **Choo, Q. L., G. Kuo, R. Ralston, A. Weiner, D. Chien, G. Van Nest, J. Han, K. Berger, K. Thudium, C. Kuo, and et al.** 1994. Vaccination of chimpanzees against infection by the hepatitis C virus. *Proc Natl Acad Sci U S A* **91**:1294-8.
15. **Choo, Q. L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**:359-62.
16. **Chowdhury, A., A. Santra, S. Chaudhuri, G. K. Dhali, S. Chaudhuri, S. G. Maity, T. N. Naik, S. K. Bhattacharya, and D. N. Mazumder.** 2003. Hepatitis C virus infection in the general population: a community-based study in West Bengal, India. *Hepatology* **37**:802-9.
17. **Cochrane, A., B. Searle, A. Hardie, R. Robertson, T. Delahooke, S. Cameron, R. S. Tedder, G. M. Dusheiko, X. De Lamballerie, and P. Simmonds.** 2002. A genetic analysis of hepatitis C virus transmission between injection drug users. *J Infect Dis* **186**:1212-21.
18. **Colina, R., D. Casane, S. Vasquez, L. Garcia-Aguirre, A. Chunga, H. Romero, B. Khan, and J. Cristina.** 2004. Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol* **85**:31-7.
19. **Cooreman, M. P., and E. M. Schoondermark-Van de Ven.** 1996. Hepatitis C virus: biological and clinical consequences of genetic heterogeneity. *Scand J Gastroenterol Suppl* **218**:106-15.
20. **Cox, A. L., T. Mosbrugger, Q. Mao, Z. Liu, X. H. Wang, H. C. Yang, J. Sidney, A. Sette, D. Pardoll, D. L. Thomas, and S. C. Ray.** 2005. Cellular immune selection with hepatitis C virus persistence in humans. *J Exp Med* **201**:1741-52.
21. **De Francesco, R., and G. Migliaccio.** 2005. Challenges and successes in developing new therapies for hepatitis C. *Nature* **436**:953-60.
22. **Dixit, V., S. Quan, P. Martin, D. Larson, M. Brezina, R. DiNello, K. Sra, J. Y. Lau, D. Chien, J. Kolberg, and et al.** 1995. Evaluation of a novel serotyping system for hepatitis C virus: strong correlation with standard genotyping methodologies. *J Clin Microbiol* **33**:2978-83.
23. **Dusheiko, G., H. Schmilovitz-Weiss, D. Brown, F. McOmish, P. L. Yap, S. Sherlock, N. McIntyre, and P. Simmonds.** 1994. Hepatitis C virus genotypes: an investigation of type-specific differences in geographic origin and disease. *Hepatology* **19**:13-8.
24. **Egger, D., B. Wolk, R. Gosert, L. Bianchi, H. E. Blum, D. Moradpour, and K. Bienz.** 2002. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol* **76**:5974-84.
25. **Farci, P., H. J. Alter, D. C. Wong, R. H. Miller, S. Govindarajan, R. Engle, M. Shapiro, and R. H. Purcell.** 1994. Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization. *Proc Natl Acad Sci U S A* **91**:7792-6.
26. **Farci, P., A. Shimoda, A. Coiana, G. Diaz, G. Peddis, J. C. Melpolder, A. Strazzer, D. Y. Chien, S. J. Munoz, A. Balestrieri, R. H. Purcell, and H. J. Alter.** 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**:339-44.

27. **Feinstone, S. M., A. Z. Kapikian, R. H. Purcell, H. J. Alter, and P. V. Holland.** 1975. Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N Engl J Med* **292**:767-70.
28. **Frank, C., M. K. Mohamed, G. T. Strickland, D. Lavanchy, R. R. Arthur, L. S. Magder, T. El Khoby, Y. Abdel-Wahab, E. S. Aly Ohn, W. Anwar, and I. Sallam.** 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* **355**:887-91.
29. **Gale, M., Jr., and E. M. Foy.** 2005. Evasion of intracellular host defence by hepatitis C virus. *Nature* **436**:939-45.
30. **Gish, R. G., K. P. Qian, S. Quan, Y. L. Xu, I. Pike, A. Polito, R. DiNello, and J. Y. Lau.** 1997. Concordance between hepatitis C virus serotyping assays. *J Viral Hepat* **4**:421-2.
31. **Gosert, R., D. Egger, V. Lohmann, R. Bartenschlager, H. E. Blum, K. Bienz, and D. Moradpour.** 2003. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J Virol* **77**:5487-92.
32. **Harada, T., N. Tautz, and H. J. Thiel.** 2000. E2-p7 region of the bovine viral diarrhoea virus polyprotein: processing and functional studies. *J Virol* **74**:9498-506.
33. **Houghton, M.** 1996. Hepatitis C viruses, p. 1035-1058. *In* B. N. Fields, Knipe, D. M. & Howley, P. M. (ed.), *Fields Virology*. Lippincott, Philadelphia.
34. **Huang, H., F. Sun, D. M. Owen, W. Li, Y. Chen, M. Gale, Jr., and J. Ye.** 2007. Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins. *Proc Natl Acad Sci U S A* **104**:5848-53.
35. **Hugle, T., F. Fehrman, E. Bieck, M. Kohara, H. G. Krausslich, C. M. Rice, H. E. Blum, and D. Moradpour.** 2001. The hepatitis C virus nonstructural protein 4B is an integral endoplasmic reticulum membrane protein. *Virology* **284**:70-81.
36. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius.** 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* **76**:4034-43.
37. **Kato, N.** 2001. Molecular virology of hepatitis C virus. *Acta Med Okayama* **55**:133-59.
38. **Kato, Y., K. Hamasaki, T. Aritomi, K. Nakao, K. Nakata, and K. Eguchi.** 1999. Most of the patients with cirrhosis in Japan die from hepatocellular carcinoma. *Oncol Rep* **6**:1273-6.
39. **Kolykhalov, A. A., S. M. Feinstone, and C. M. Rice.** 1996. Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol* **70**:3363-71.
40. **Koziel, M. J., D. Dudley, J. T. Wong, J. Dienstag, M. Houghton, R. Ralston, and B. D. Walker.** 1992. Intrahepatic cytotoxic T lymphocytes specific for hepatitis C virus in persons with chronic hepatitis. *J Immunol* **149**:3339-44.
41. **Kuiken, C., P. Hraber, J. Thurmond, and K. Yusim.** 2008. The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res* **36**:D512-6.
42. **Kumar, U., J. Brown, J. Monjardino, and H. C. Thomas.** 1993. Sequence variation in the large envelope glycoprotein (E2/NS1) of hepatitis C virus during chronic infection. *J Infect Dis* **167**:726-30.

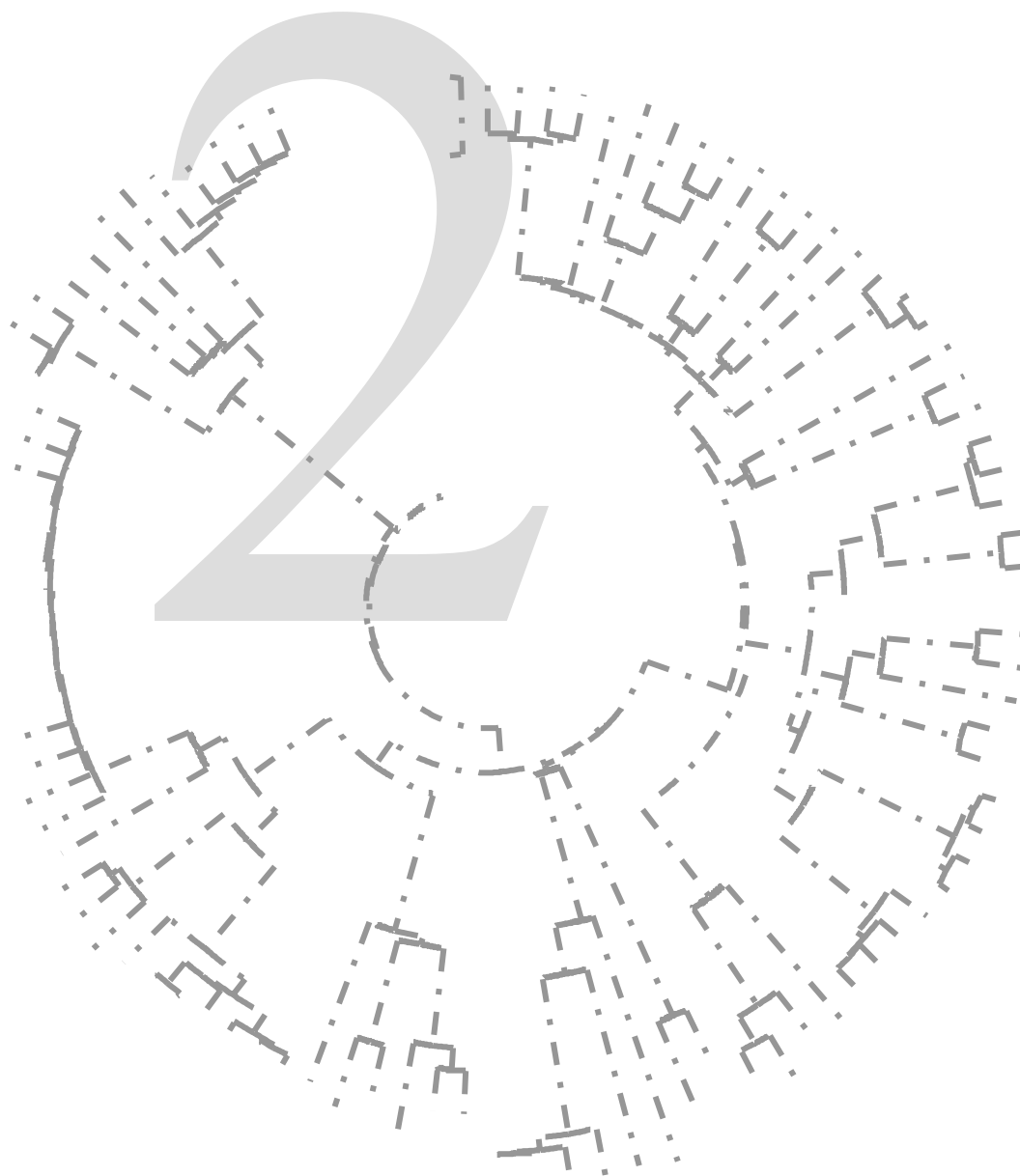
43. **Lai, M. E., A. P. Mazzoleni, F. Argioli, S. De Virgilis, A. Balestrieri, R. H. Purcell, A. Cao, and P. Farci.** 1994. Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. *Lancet* **343**:388-90.
44. **Lai, M. M.** 1992. RNA recombination in animal and plant viruses. *Microbiol Rev* **56**:61-79.
45. **Lerat, H., M. Honda, M. R. Beard, K. Loesch, J. Sun, Y. Yang, M. Okuda, R. Gosert, S. Y. Xiao, S. A. Weinman, and S. M. Lemon.** 2002. Steatosis and liver cancer in transgenic mice expressing the structural and nonstructural proteins of hepatitis C virus. *Gastroenterology* **122**:352-65.
46. **Li, X., L. J. Jeffers, L. Shao, K. R. Reddy, M. de Medina, J. Scheffel, B. Moore, and E. R. Schiff.** 1995. Identification of hepatitis C virus by immunoelectron microscopy. *J Viral Hepat* **2**:227-34.
47. **Lindenbach, B. D., and C. M. Rice.** 2005. Unravelling hepatitis C virus replication from genome to function. *Nature* **436**:933-8.
48. **Lohmann, V., J. O. Koch, and R. Bartenschlager.** 1996. Processing pathways of the hepatitis C virus proteins. *J Hepatol* **24**:11-9.
49. **Lohmann, V., F. Korner, J. Koch, U. Herian, L. Theilmann, and R. Bartenschlager.** 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**:110-3.
50. **Maclennan, S., J. A. Barbara, P. Hewitt, C. Moore, and M. Contreras.** 1992. Screening blood donations for HCV. *Lancet* **339**:131-2.
51. **Makuwa, M., S. Souquiere, P. Telfer, E. Leroy, O. Bourry, P. Rouquet, S. Clifford, E. J. Wickings, P. Roques, and F. Simon.** 2003. Occurrence of hepatitis viruses in wild-born non-human primates: a 3 year (1998-2001) epidemiological survey in Gabon. *J Med Primatol* **32**:307-14.
52. **Marcellin, P.** 1999. Hepatitis C : the clinical spectrum of the disease. *J Hepatol* **31 Suppl 1**:9-16.
53. **Martell, M., J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez.** 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* **66**:3225-9.
54. **McHutchison, J. G., S. C. Gordon, E. R. Schiff, M. L. Shiffman, W. M. Lee, V. K. Rustgi, Z. D. Goodman, M. H. Ling, S. Cort, and J. K. Albrecht.** 1998. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *N Engl J Med* **339**:1485-92.
55. **Meylan, E., J. Curran, K. Hofmann, D. Moradpour, M. Binder, R. Bartenschlager, and J. Tschopp.** 2005. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature* **437**:1167-72.
56. **Michielsen, P. P., S. M. Francque, and J. L. van Dongen.** 2005. Viral hepatitis and hepatocellular carcinoma. *World J Surg Oncol* **3**:27.
57. **Moradpour, D., F. Penin, and C. M. Rice.** 2007. Replication of hepatitis C virus. *Nat Rev Microbiol* **5**:453-63.
58. **Mujeeb, S. A., S. Shahab, and A. A. Hyder.** 2000. Geographical display of health information: study of hepatitis C infection in Karachi, Pakistan. *Public Health* **114**:413-5.

59. **Nakano, T., L. Lu, P. Liu, and O. G. Pybus.** 2004. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J Infect Dis* **190**:1098-108.
60. **Ndjomou, J., O. G. Pybus, and B. Matz.** 2003. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *J Gen Virol* **84**:2333-41.
61. **Neumann, A. U., N. P. Lam, H. Dahari, D. R. Gretch, T. E. Wiley, T. J. Layden, and A. S. Perelson.** 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* **282**:103-7.
62. **O'Connor, D. H., A. B. McDermott, K. C. Krebs, E. J. Dodds, J. E. Miller, E. J. Gonzalez, T. J. Jacoby, L. Yant, H. Piontkivska, R. Pantophlet, D. R. Burton, W. M. Rehrauer, N. Wilson, A. L. Hughes, and D. I. Watkins.** 2004. A dominant role for CD8⁺-T-lymphocyte selection in simian immuno-deficiency virus sequence variation. *J Virol* **78**:14012-22.
63. **Okamoto, H., M. Kojima, S. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiro.** 1992. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* **190**:894-9.
64. **Okamoto, H., Y. Sugiyama, S. Okada, K. Kurai, Y. Akahane, Y. Sugai, T. Tanaka, K. Sato, F. Tsuda, Y. Miyakawa, and et al.** 1992. Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *J Gen Virol* **73 (Pt 3)**:673-9.
65. **Pavlovic, D., D. C. Neville, O. Argaud, B. Blumberg, R. A. Dwek, W. B. Fischer, and N. Zitzmann.** 2003. The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc Natl Acad Sci U S A* **100**:6104-8.
66. **Pawlotsky, J. M., L. Tsakiris, F. Roudot-Thoraval, C. Pellet, L. Stuyver, J. Duval, and D. Dhumeaux.** 1995. Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. *J Infect Dis* **171**:1607-10.
67. **Penin, F., J. Dubuisson, F. A. Rey, D. Moradpour, and J. M. Pawlotsky.** 2004. Structural biology of hepatitis C virus. *Hepatology* **39**:5-19.
68. **Pietschmann, T., A. Kaul, G. Koutsoudakis, A. Shavinskaya, S. Kallis, E. Steinmann, K. Abid, F. Negro, M. Dreux, F. L. Cosset, and R. Bartenschlager.** 2006. Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc Natl Acad Sci U S A* **103**:7408-13.
69. **Prati, D.** 2006. Transmission of hepatitis C virus by blood transfusions and other medical procedures: a global review. *J Hepatol* **45**:607-16.
70. **Pybus, O. G., M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey.** 2001. The epidemic behavior of the hepatitis C virus. *Science* **292**:2323-5.
71. **Pybus, O. G., A. J. Drummond, T. Nakano, B. H. Robertson, and A. Rambaut.** 2003. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* **20**:381-7.
72. **Ray, S. C., R. R. Arthur, A. Carella, J. Bukh, and D. L. Thomas.** 2000. Genetic epidemiology of hepatitis C virus throughout egypt. *J Infect Dis* **182**:698-707.

73. **Ray, S. C., L. Fanning, X. H. Wang, D. M. Netski, E. Kenny-Walsh, and D. L. Thomas.** 2005. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J Exp Med* **201**:1753-9.
74. **Robertson, B., G. Myers, C. Howard, T. Brettin, J. Bukh, B. Gaschen, T. Gojobori, G. Maertens, M. Mizokami, O. Nainan, S. Netesov, K. Nishioka, T. Shin i, P. Simmonds, D. Smith, L. Stuyver, and A. Weiner.** 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**:2493-503.
75. **Sanchez-Quijano, A., J. Andreu, F. Gavilan, F. Luque, M. A. Abad, B. Soto, J. Munoz, J. M. Aznar, M. Leal, and E. Lissen.** 1995. Influence of human immunodeficiency virus type 1 infection on the natural course of chronic parenterally acquired hepatitis C. *Eur J Clin Microbiol Infect Dis* **14**:949-53.
76. **Serebrov, V., and A. M. Pyle.** 2004. Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature* **430**:476-80.
77. **Shepard, C. W., L. Finelli, and M. J. Alter.** 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* **5**:558-67.
78. **Silini, E., F. Bono, A. Cividini, A. Cerino, S. Bruno, S. Rossi, G. Belloni, B. Brugnetti, E. Civardi, L. Salvaneschi, and et al.** 1995. Differential distribution of hepatitis C virus genotypes in patients with and without liver function abnormalities. *Hepatology* **21**:285-90.
79. **Simmonds, P.** 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol* **85**:3173-88.
80. **Simmonds, P.** 2001. Reconstructing the origins of human hepatitis viruses. *Philos Trans R Soc Lond B Biol Sci* **356**:1013-26.
81. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-73.
82. **Simmonds, P., J. Mellor, T. Sakuldamrongpanich, C. Nuchaprayoon, S. Tanprasert, E. C. Holmes, and D. B. Smith.** 1996. Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *J Gen Virol* **77 (Pt 12)**:3013-24.
83. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 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.
84. **Smith, D. B., J. Mellor, L. M. Jarvis, F. Davidson, J. Kolberg, M. Urdea, P. L. Yap, and P. Simmonds.** 1995. Variation of the hepatitis C virus 5' non-coding region: implications for secondary structure, virus detection and typing. The International HCV Collaborative Study Group. *J Gen Virol* **76 (Pt 7)**:1749-61.
85. **Smith, D. B., S. Pathirana, F. Davidson, E. Lawlor, J. Power, P. L. Yap, and P. Simmonds.** 1997. The origin of hepatitis C virus genotypes. *J Gen Virol* **78 (Pt 2)**:321-8.

86. **Smith, D. B., and P. Simmonds.** 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**:238-46.
87. **Smith, D. B., and P. Simmonds.** 1997. Review: molecular epidemiology of hepatitis C virus. *J Gastroenterol Hepatol* **12**:522-7.
88. **Songsivilai, S., D. Kanistanon, and T. Dharakul.** 1998. A serotyping assay for hepatitis C virus in Southeast Asia. *Clin Diagn Lab Immunol* **5**:737-9.
89. **Soto, B., A. Sanchez-Quijano, L. Rodrigo, J. A. del Olmo, M. Garcia-Bengochea, J. Hernandez-Quero, C. Rey, M. A. Abad, M. Rodriguez, M. Sales Gilabert, F. Gonzalez, P. Miron, A. Caruz, F. Relimpio, R. Torronteras, M. Leal, and E. Lissen.** 1997. Human immunodeficiency virus infection modifies the natural history of chronic parenterally-acquired hepatitis C with an unusually rapid progression to cirrhosis. *J Hepatol* **26**:1-5.
90. **Stuyver, L., R. Rossau, A. Wyseur, M. Duhamel, B. Vanderborght, H. Van Heuverswyn, and G. Maertens.** 1993. Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *J Gen Virol* **74 (Pt 6)**:1093-102.
91. **Stuyver, L., A. Wyseur, W. van Arnhem, F. Hernandez, and G. Maertens.** 1996. Second-generation line probe assay for hepatitis C virus genotyping. *J Clin Microbiol* **34**:2259-66.
92. **Sulaiman, H. A., Julitasari, A. Sie, M. Rustam, W. Melani, A. Corwin, and G. B. Jennings.** 1995. Prevalence of hepatitis B and C viruses in healthy Indonesian blood donors. *Trans R Soc Trop Med Hyg* **89**:167-70.
93. **Tan, S. L., and M. G. Katze.** 2001. How hepatitis C virus counteracts the interferon response: the jury is still out on NS5A. *Virology* **284**:1-12.
94. **Tanaka, T., N. Kato, M. J. Cho, K. Sugiyama, and K. Shimotohno.** 1996. Structure of the 3' terminus of the hepatitis C virus genome. *J Virol* **70**:3307-12.
95. **Tanaka, T., N. Kato, M. Nakagawa, Y. Ootsuyama, M. J. Cho, T. Nakazawa, M. Hijikata, Y. Ishimura, and K. Shimotohno.** 1992. Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals. *Virus Res* **23**:39-53.
96. **Tellinghuisen, T. L., and C. M. Rice.** 2002. Interaction between hepatitis C virus proteins and host cell factors. *Curr Opin Microbiol* **5**:419-27.
97. **Tokita, H., H. Okamoto, H. Iizuka, J. Kishimoto, F. Tsuda, L. A. Lesmana, Y. Miyakawa, and M. Mayumi.** 1996. Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups. *J Gen Virol* **77 (Pt 2)**:293-301.
98. **Tokita, H., H. Okamoto, F. Tsuda, P. Song, S. Nakata, T. Chosa, H. Iizuka, S. Mishiro, Y. Miyakawa, and M. Mayumi.** 1994. Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *Proc Natl Acad Sci U S A* **91**:11022-6.
99. **Tsukiyama-Kohara, K., N. Iizuka, M. Kohara, and A. Nomoto.** 1992. Internal ribosome entry site within hepatitis C virus RNA. *J Virol* **66**:1476-83.
100. **Twiddy, S. S., and E. C. Holmes.** 2003. The extent of homologous recombination in members of the genus *Flavivirus*. *J Gen Virol* **84**:429-40.
101. **Vento, S., T. Garofano, C. Renzini, F. Cainelli, F. Casali, G. Ghironzi, T. Ferraro, and E. Concia.** 1998. Fulminant hepatitis associated with hepatitis A

- virus superinfection in patients with chronic hepatitis C. *N Engl J Med* **338**:286-90.
102. **Wakita, T., T. Pietschmann, T. Kato, T. Date, M. Miyamoto, Z. Zhao, K. Murthy, A. Habermann, H. G. Krausslich, M. Mizokami, R. Bartenschlager, and T. J. Liang.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**:791-6.
 103. **Wasley, A., and M. J. Alter.** 2000. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin Liver Dis* **20**:1-16.
 104. **Weiner, A. J., H. M. Geysen, C. Christopherson, J. E. Hall, T. J. Mason, G. Saracco, F. Bonino, K. Crawford, C. D. Marion, K. A. Crawford, and et al.** 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc Natl Acad Sci U S A* **89**:3468-72.
 105. **WHO.** 2000. Hepatitis C--global prevalence (update). *Wkly Epidemiol Rec* **75**:18-19.
 106. **WHO.** 1997. Hepatitis C: global prevalence. *Wkly Epidemiol Rec* **72**:341-4.
 107. **You, S., D. D. Stump, A. D. Branch, and C. M. Rice.** 2004. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J Virol* **78**:1352-66.
 108. **Zarski, J. P., B. Bohn, A. Bastie, J. M. Pawlotsky, M. Baud, F. Bost-Bezeaux, J. Tran van Nhieu, J. M. Seigneurin, C. Buffet, and D. Dhumeaux.** 1998. Characteristics of patients with dual infection by hepatitis B and C viruses. *J Hepatol* **28**:27-33.
 109. **Zein, N. N., J. Rakela, E. L. Krawitt, K. R. Reddy, T. Tominaga, and D. H. Persing.** 1996. Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. Collaborative Study Group. *Ann Intern Med* **125**:634-9.
 110. **Zylberberg, H., M. L. Chaix, and C. Brechot.** 2000. Infection with hepatitis C virus genotype 4 is associated with a poor response to interferon-alpha. *Ann Intern Med* **132**:845-6.



CHAPTER 2

Genotyping Hepatitis C Viruses from South East Asia by a Novel Line Probe Assay That Simultaneously Detects Core and 5' Untranslated Regions

Suwanna Noppornpanth,¹ Erwin Sablon,² Kathy De Nys,² Truong Xuan Lien,³ Jan Brouwer,¹ Marianne Van Brussel,² Saskia L. Smits,¹ Yong Poovorawan,⁴ Albert D. M. E. Osterhaus¹ and Bart L. Haagmans¹

*Department of Virology, Erasmus Medical Center, Rotterdam, The Netherlands¹
Innogenetics, N.V., Gent, Belgium²*

Department of Biological Analysis, Pasteur Institute Ho Chi Minh City, Ho Chi Minh City, Vietnam³

Center of Excellence in Viral Hepatitis Research, Department of Pediatrics, Chulalongkorn Hospital, Bangkok, Thailand.⁴

Journal of Clinical Microbiology, 44 (2006), p. 3969-74.

ABSTRACT

Hepatitis C viruses (HCVs) display a high level of sequence diversity and are currently divided into six genotypes. A line probe assay (LiPA), which targets the 5'untranslated region (5'UTR) of the HCV genome, is widely used for genotyping. However, this assay cannot distinguish many genotype 6 subtypes from genotype 1 due to high sequence similarity in the 5'UTR. We investigated the accuracy of a new generation LiPA (VERSANT HCV Genotype 2.0 assay), in which genotyping is based on 5'UTR and core sequences, by testing 75 selected HCV RNA positive sera from South East Asia (Vietnam and Thailand). For comparison, sera were tested on the 5'UTR based VERSANT HCV Genotype assay (LiPA) and processed for sequence analysis of the 5'UTR-to-core and NS5b regions as well. Phylogenetic analysis of both regions revealed the presence of genotype 1, 2, 3 and 6 viruses. Using the new LiPA assay, genotype 6c-6l and 1a/b samples were more accurately genotyped than with the previous test only targeting the 5'UTR (96% versus 71% respectively). These results indicate that the VERSANT HCV Genotype 2.0 assay is able to discriminate genotype 6c-6l from genotype 1, and allows a more accurate identification of genotype 1a from 1b by using the genotype-specific core information.

INTRODUCTION

Hepatitis C virus (HCV), an enveloped positive-stranded RNA virus of the family *Flaviviridae*, is recognized as a major cause of chronic liver disease. Because of its high genetic heterogeneity, HCV has been classified into 6 genotypes and a huge number of subtypes (18, 20). Genotypes 1, 2, and 3 are widely distributed around the world while genotypes 4 and 5 have been identified mainly in Africa (22). Genotype 6 was found locally in South East Asia (SEA) (11, 26, 27, 28).

Genotype identification is clinically important for prediction of responses to, and in determining the duration of, antiviral therapy (30). This is illustrated by the fact that genotypes 1 and 4 are more resistant to treatment with pegylated interferon- α and ribavirin than genotypes 2 and 3 (9, 31). Moreover, it has been suggested that patients with chronic HCV genotype 1b infection show more severe liver disease than patients infected with other genotypes (17). Nowadays, most treatment protocols require preceding genotype information of HCV infected patients.

A variety of technologies has been developed for HCV genotype determination. Most of these assays rely on the amplification of short HCV RNA regions from clinical specimens, followed by a type-specific assay, such as line probe reverse hybridization (23, 25), restriction fragment length polymorphism analysis (1), or sequence analysis (20, 29). Almost all available commercial assays target the 5'UTR, because the highly conserved sequences of this region are most suitable for RT-PCR amplification.

The VERSANT HCV Genotype assay (LiPA) is one of the most widely used methods for HCV genotyping. In this assay the 5'UTR of HCV is amplified with biotinylated primers, after which the PCR product is hybridized to a membrane impregnated with genotype-specific probes and detected with streptavidine linked to a colorimetric detector (24). Despite the high conservation of the 5'UTR, genotype determination of HCV based on the 5'UTR is accurate for most genotypes (8, 21, 29). However, it has been noted that methods which are based on the use of the 5'UTR falsely identify genotypes 6c-6l from SEA as genotype 1, which is also the case in the VERSANT HCV Genotype assay (3, 24). Moreover, this assay is unable to distinguish genotype 1a from 1b in 5-10% of the cases (2, 21). Therefore, it has been recommended

to use other coding regions of the HCV genome (e.g. core, E1, and/or NS5b) for genotype identification (18).

A new generation of the line probe assay (VERSANT HCV Genotype 2.0 assay) was recently developed, which uses core sequence information, in addition to 5'UTR, to improve the accuracy of HCV genotyping. After amplification of the 5'UTR-to-core region and hybridization to type-specific probes of the 5'UTR and core, it is possible to distinguish between genotypes 1a, 1b and 6c-6l. In the present study, we selected 75 sera of HCV RNA-positive blood donors collected from Bangkok, Thailand and Ho Chi Minh City, Vietnam in order to test the accuracy of the new VERSANT HCV Genotype 2.0 assay. The data were compared to results obtained with the previous VERSANT HCV Genotype method and sequence analysis. We show that the VERSANT HCV Genotype 2.0 assay is able to discriminate genotypes 6c-6l from genotype 1, and allows a more accurate identification of genotype 1a from 1b by using the genotype-specific core information.

MATERIALS AND METHODS

Samples. A total of 152 HCV-positive plasma samples of blood donors from Ho Chi Minh City, Vietnam and Bangkok, Thailand were obtained during 2000-2002. From this collection of samples, 75 were selected based on an initial characterization to include preferentially genotype 6 viruses (12, 26). Samples were kept at -80°C until further analysis.

RNA isolation and RT-PCR. Viral RNA was isolated from 200 μl serum or EDTA-plasma by the High Pure RNA isolation Kit (Roche, Diagnostics GmbH) and eluted with 50 μl water. cDNA was synthesized using 10 μl extracted RNA, 2 μl random hexamer primers (150 $\mu\text{g}/\text{ml}$; Boehringer Mannheim) and MMLV Reverse transcriptase (Invitrogen) according to the instructions of the manufacturers. The 5'UTR-core sequences (nucleotides 47 to 695, according to the numbering system for reference strain HCV-H Genbank accession number M67463) were amplified by nested PCR with primers 16 (5'-GRGGCGACTCCACCAT-3'), 410 and primers s17 and 951, as described previously (10). NS5b sequences (nucleotides 8283 to 8624, according to the numbering system for HCV-H strain) were amplified by nested PCR with primers Pr3,

Pr4, Pr1, and Pr2 using conditions described in previous reports (16) with modifications. Briefly, in the first PCR, performed with Pr3 and Pr4, 1.5 mM MgCl₂ was used in the reaction mixture, and the following thermal profile: initial denaturation at 94°C for 90 s; 5 cycles of denaturation at 94°C for 30s, annealing at 64°C for 45 s and elongation at 72°C for 1 min, followed immediately by 30 cycles at 94°C for 30 s, 64°C with a drop of -0.5°C between each cycle for 45s and elongation at 72°C for 1 min. The last 5 cycles were performed at 94°C for 30 s, 48°C for 45 s and 72°C for 1 min. A final elongation at 72°C for 10 min was also included. The nested PCR, performed with primer Pr1 and Pr2, was carried out on 2 µl of the first PCR product with the following thermal profile: 95°C for 7 min; 50 cycles at 95°C for 30 s, 63°C for 30 s and 72°C for 30 s, with a final elongation step at 72°C for 10 min. The amplified products were gel-purified and sequenced.

DNA sequencing and phylogenetic analysis. Sequence reactions on PCR products were performed using BigDye Terminator Cycle sequencing Ready Reaction kit (Applied Biosystems) and an ABI prism 3100 Autosequencer (Applied Biosystems). Multiple sequence alignments were generated with BioEdit version 7.0.1. Phylogenetic analyses were conducted using the Tajima-Nei model with rate heterogeneity parameter and Neighbor-joining method. Confidence values were calculated by bootstrap analysis (1000 replicates) and consensus trees were produced by MEGA version 3.0 (7).

Reference strains used in this study were obtained from Genbank: genotype 1a (M67463), 1b (AB016785), 1c (AY051292), 2a (AB047639), 2b (AB030907), 2c (D50409), 2k (AB031663), 3a (AF046866), 3b (D49374), 3k (10a, D63821), 4a (D45193, Y11604), 5a (D50466, Y13184), 6a (D88469, AY859526), 6b (D84262), 6c (7d, D37843, D37885), 6d (7b, D84263), 6e (7a, D31971, D30397), 6f (7c, D38078, D38078), 6g (11a, D63822), 6h (9a, D84265), 6i (9b, D37850, D37864), 6j (9c, D37848, D37862), 6k (8b, D84264), 6l (8a, D88470, D87357), 6n (DQ278894).

INNO-LiPA line probe assay. Sera samples were extracted with the Viral DSP kit (Qiagen) according to the manufacturer's instructions, with elution volume of 60 µl. The cDNA synthesis and amplifications for VERSANT HCV Genotype assay (Bayer HealthCare, manufactured by Innogenetics, Ghent, Belgium) were performed according to the manufactures instructions. For the VERSANT HCV Genotype 2.0 assay (Bayer HealthCare, manufactured by Innogenetics, Ghent, Belgium) a multiplex RT-PCR was

performed according to the manufactures instructions on the extracts to amplify the 5'UTR and core regions of the HCV genome, utilizing primers HCVPr95 AT&AC, HCVPr96b, HCVPr769b, HCVPr822b, generating two distinct PCR products from 5'UTR and core regions, respectively. Briefly, master mixes were prepared using these primers, buffer, enzymes, and deoxynucleoside triphosphates (dNTPs), including deoxyuridine triphosphate (dUTP), from the VERSANT HCV Amplification 2.0 assay. Twenty microliters of RNA extract was added to the master mix and treated with uracil-*N*-glycosylase to prevent contamination by amplification product from previous PCRs. Subsequently, the mix was incubated for 30 minutes at 50°C and 15 minutes at 95°C for RT and PCR activation respectively, immediately followed by 40 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 15 s. After amplification, the PCR products were immediately stored at -20°C, to prevent residual uracil-*N*-glycosylase activity. Hybridization and color development steps were performed in the Auto-LiPA48 instrument (Innogenetics, Ghent, Belgium), according to the manufacturer's instructions and were not different for the two assays. Strip patterns obtained from both assays were interpreted following the manufacturer's interpretation chart. HCV genotyping results from the LiPA assays were compared to the gold standard phylogenetic analysis of core and NS5b sequences. The sensitivity of the assay was determined by probit analysis to 2106 IU/ml.

Nucleotide sequence accession numbers. New sequences reported in this study have been submitted to Genbank database and assigned accession numbers DQ155445, DQ155447-455, DQ155457, DQ15460-463, DQ15465-474, DQ155477-493, DQ155495-500, DQ640336-361 for 5'UTR-to-core phylogenetic analysis. Sequences accession numbers DQ155503, DQ155505-513, DQ155515, DQ155518-521, DQ155523-532, DQ155535-551, DQ155553-558, DQ640362-386 were used for NS5b sequences analysis.

RESULTS

Determination of HCV genotype by sequence analysis. To investigate the accuracy of the newly developed VERSANT HCV Genotype 2.0 assay for genotype 6 identification, we selected a set of samples from SEA HCV positive blood donors to encompass mainly genotype 1 and 6 viruses. A total of 75 samples, 50 from Vietnam and 25 from Thailand,

were included in this study. Viral RNA served as template for RT-PCR amplification of 5'UTR-core (nucleotides 47-695) and NS5b (nucleotides 8283-8624) sequences. Amplicons were sequenced twice in both orientations. Degenerate sequences were preserved in order to determine the specificity of the line probe hybridization assays (see below). To determine the HCV genotype, phylograms were constructed with reference strains for 5'UTR-core and NS5b regions (Fig.1 and data not shown). Only samples which showed genotype consistency of both regions were selected (n = 73) and 2 discordant samples (D3 and D54) which turned out to be a recombinant virus and a mixed infection were characterized by full-length genome analysis as described elsewhere (12). As evident from the 5'UTR-core phylogenetic analysis (Fig. 1, Table 1), our sample collection contained genotype 1 (n = 29), 2 (n = 4), 3 (n = 7) and 6 (n = 33) viruses.

Determination of HCV genotype by the VERSANT HCV Genotype (5'UTR) assay.

To examine the accuracy of HCV genotyping by the line probe hybridization assay based on the 5'UTR sequence, we tested all selected samples using the VERSANT HCV Genotype assay (Fig. 2A; Table 1). This assay correctly determined the *genotype* of 71% (52/73) of the samples as compared to sequence analysis. From these identified *genotypes*, 15% (8/52) were *subtype* indeterminate of which 4 were of subtype 1a, 2 of subtype 1b, 1 of subtype 2a and 1 of subtype 3b. Only one sample (2%) of genotype 1a was incorrectly subtyped as genotype 1b (Table 1). All 21 samples that could not be genotyped correctly are genotypes 6c-6l that were typed to genotype 1. The 5'UTR sequences of the 6c-6l samples are very similar to the genotype 1 5'UTR sequences, especially in the regions where probes 3, 4, and 6 are located (24, Fig. 3), which explains the mistyping of the genotypes 6c-6l in this assay.

Determination of HCV genotype by the VERSANT HCV Genotype 2.0 (5'UTR-core) assay.

The new HCV LiPA assay consists of 25 lines in total (Fig. 2B). Lines 1 to 21 are identical to the strip configuration of the VERSANT HCV Genotype assay (Fig. 2A), which is based on the 5'UTR sequences. Another four lines specific for the core region are used for control core amplification, genotype 6c-6l, 1a, and 1b, respectively. These core specific lines are only to be considered when 5'UTR line patterns refer to genotype 1.

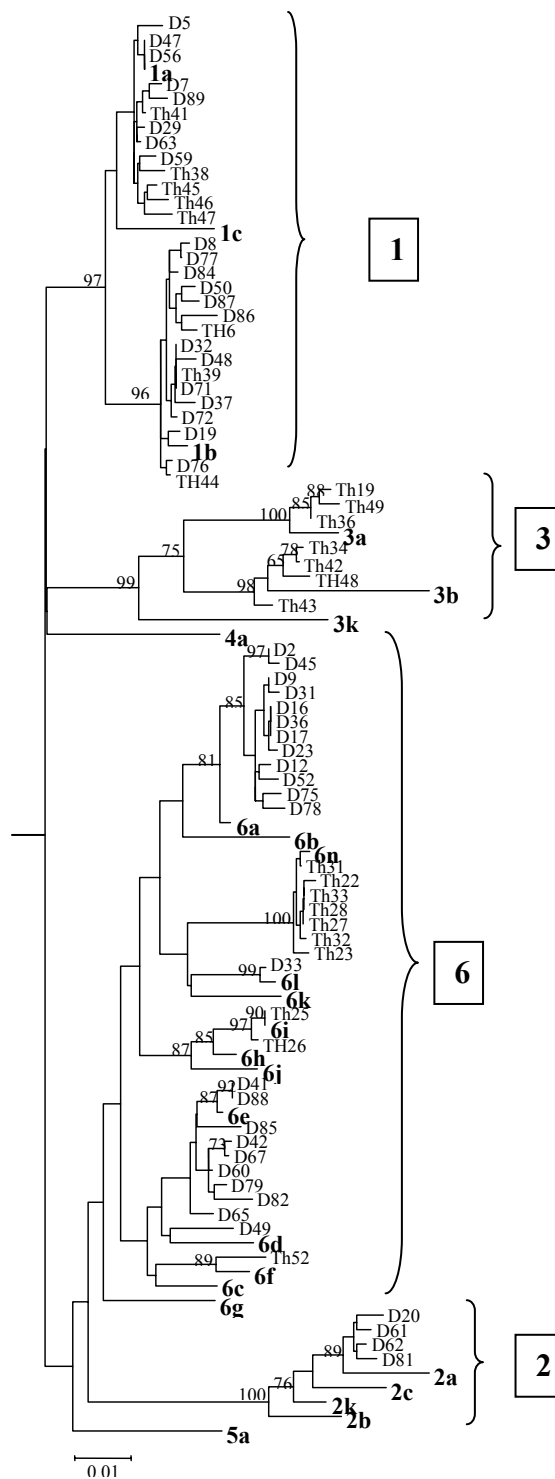


Figure 1 The midpoint rooted neighbor-joining tree of the 5'UTR-to-core region among selected HCV field viruses from SEA and prototype strains. The tree was constructed for nucleotides 47 to 695 of the HCV genome (numbering according to reference strain HCV-H Genbank accession number M67463) by using the Tajima-Nei model with rate heterogeneity parameter. Confidence values (>70%) calculated by bootstrap analysis (1,000 replicates) are indicated at the major branching points. Branch lengths are drawn to scale. The prototype HCV strains obtained from Genbank are indicated in bold.

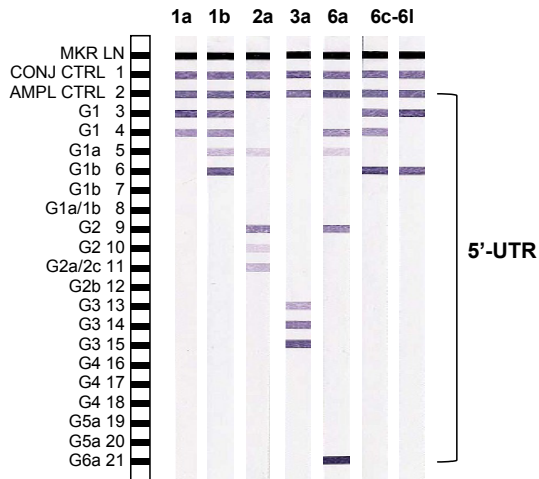
TABLE 1. Comparison of HCV genotyping results obtained by phylogenetic analysis, the VERSANT HCV Genotype assay (5'UTR), and the VERSANT HCV Genotype 2.0 assay (5'UTR-core) for 73 specimens obtained from SEA.

Genotype	No. of HCV genotyped samples (type or subtype) as determined by ^a			
	Phylogenetic analysis		VERSANT HCV (5'UTR)	VERSANT HCV2.0 (5'UTR/Core)
	5'UTR/core	NS5B		
Genotype 1				
Subtype 1a	13	13	6 (1a) 2 (1a or 1b) 1 (1b) 4 (1)	11 (1a) 2 (1)*
Subtype 1b	16	16	14 (1b) 2 (1)	14 (1b) 2 (NT)*
Genotype 2				
Subtype 2a	4	4	3 (2a or 2c) 1 (2)	4 (2a or 2c)
Genotype 3				
Subtype 3a	3	3	3 (3a)	3 (3a)
Subtype 3b	4	4	3 (3b) 1 (3)	3 (3b) 1 (3)*
Genotype 6				
Subtype 6a	12	12	12 (6a)	12 (6a or 6b)
Subtype 6c to 6n	21	21	19 (1b) 2 (1)	20 (6c to 6l) 1 (NT)*

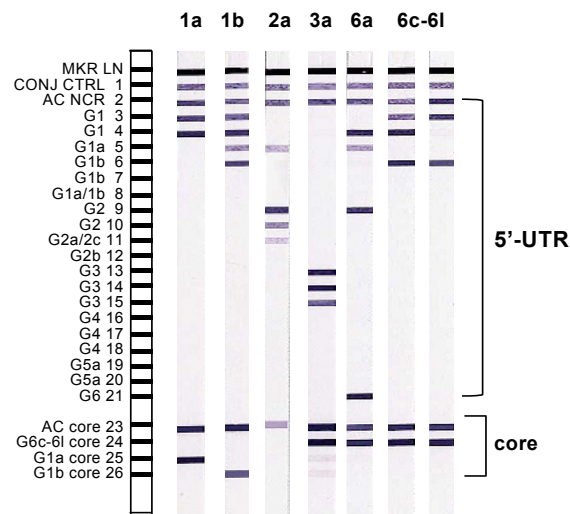
^aNT, Nontypeable. *, INNO-LiPA line patterns shown in Fig. 2C

Figure 2 (A and B) Comparison of INNO-LiPA strip patterns obtained from VERSANT HCV Genotype assay (5'UTR lines) (A) and the new VERSANT HCV Genotype 2.0 assay (5'UTR/core lines) (B). Both assays were tested with HCV genotype 1a, 1b, 2a, 3a, 6a and 6c-6l samples. Type-specific line numbers and interpreted genotypes of 5'UTR and core regions are indicated on the left side of each panel. (C) Patterns of the VERSANT HCV Genotype 2.0 assay obtained from subtype-indeterminate samples (from Th41, D89 [genotype 1a], Th48 [genotype 3b] and nontypeable samples (from Th6, Th44 [genotype 1b], Th52 [genotype 6f]. Sample codes and specific probe numbers are indicated. MKR, marker line; CONJ CTRL, conjugate control; AMPL CTRL, amplification control; AC NCR, amplification control of noncoding region; AC CORE, amplification control of core.

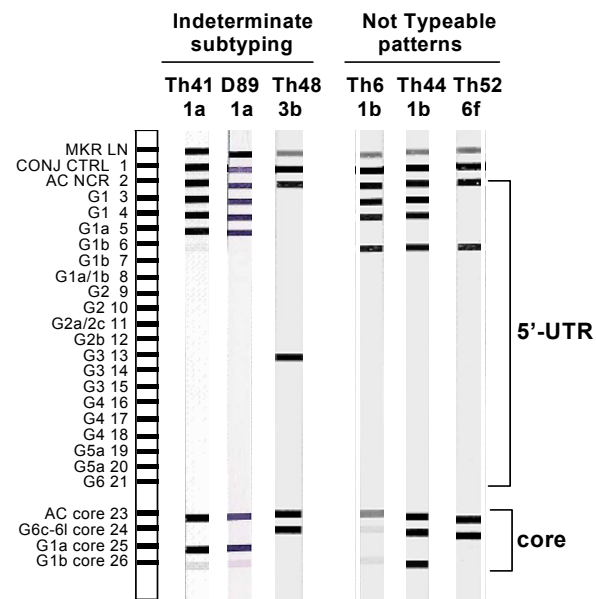
2A)



2B)



2C)



Upon analysis of the same sample set as described above, the new VERSANT HCV Genotype 2.0 assay correctly classified the HCV genotype in 96% (70/73) of the samples. All genotype 6 samples could be distinguished from genotype 1 viruses. This result is in agreement with sequence comparisons of HCV core region obtained from these samples, showing differences between genotype 1 and 6 viruses (Fig. 3). Incorporation of core probes in the assay therefore explains the more accurate HCV genotype 6 classification results with the new VERSANT HCV Genotype 2.0 assay (Table 1). Only three samples could not be subtyped; the subtype 1a samples Th41 and D89, showed strong binding of core specific line 25 (subtype 1a), but also a faint band at the position of line 26 (subtype 1b), while the subtype 3b sample Th 48 showed only UTR specific line 13 (Fig. 2C).

Three samples (4%) could not be genotyped; sample Th52 (genotype 6 by sequence analysis) showed binding to line 24 indicating a subtype 6 virus, whereas there was no binding to the 5'UTR lines 3 and 4 (Fig. 2C). As shown in figure 3, this result can be explained by a single nucleotide substitution in the Th52 core sequence at a position where probes 3 and 4 bind (nucleotide 183, C→ T and nucleotide 215, A→ G, respectively). On the other hand, samples Th6 and Th44 (subtype 1b by sequence analysis) bound to core lines 24 and 26 (genotype 6 and 1b, Fig. 2C). To test the possibility of double infection in these samples, RT-PCR products of UTR-core regions of Th6 and Th44 samples were cloned and at least 25 clones per sample were sequenced. No evidence of a mixed genotype infection could be obtained in the samples tested, although the presence of a co-infecting strain as a minor population cannot entirely be ruled out.

DISCUSSION

Clinical studies have shown that the HCV genotype is an important predictor for efficiency of antiviral treatment. Therefore, many efforts have been made to develop assays for HCV genotyping and several commercial tests are available (13). The gold standard for determination of HCV genotype is sequence analysis of phylogenetically informative coding regions of the HCV genome and comparison to consensus sequences of known genotypes. However, sequence analysis of amplicons is laborious and is not accurate in identifying mixed infections. Moreover, as the informative sequences on the

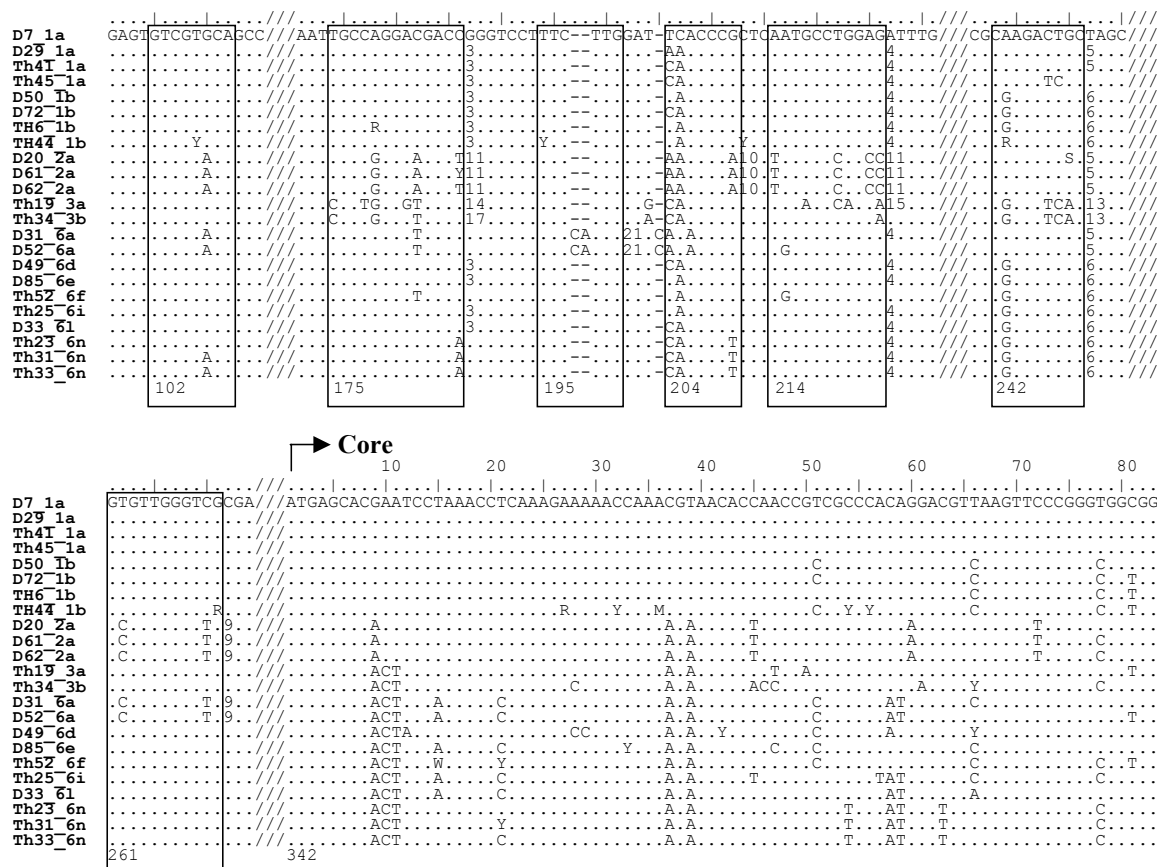


Figure 3 Alignment of the 5'UTR and core sequences of a subset of samples obtained from SEA. Sample codes and genotypes are indicated in the left-hand column. Residues identical to the major sequence are indicated by dots. The positions of specific probes of the VERSANT HCV Genotype assay (5'UTR) are boxed. The first nucleotide position of each box is indicated at the bottom left corner. The LiPA line number outside the box indicates the predicted reactivity of each isolate in the sequence alignment. The core sequence alignment starts with the ATG initiation codon (positions 342 to 345 of the HCV genome; the numbering system is as described for Fig.1). Slashes indicated discontinuous sequences within the 5'UTR.

genome are also regions with a relatively high heterogeneity, primers may fail to anneal, resulting in assay failures.

Hybridization of HCV 5'UTR sequences to genotype-specific probes with the VERSANT HCV Genotype assay has so far been the keystone for HCV genotyping, because the high degree of sequence conservation makes this region ideal for RT-PCR amplification. However, this region contains lower phylogenetic information than other genomic sequences such as core and NS5b (15). Not surprisingly, its ability to discriminate between HCV genotypes was challenged by the high diversity of HCV sequences, especially by the genotypes 6c-6l from SEA (19, 21, 24, this study). Because of the high homology of the 5'UTR between genotype 1 and genotype 6c-6l, the latter

are claimed to be genotype 1 by the assay (24, Fig. 2, Table 1). Moreover, it has been reported that the VERSANT HCV Genotype assay could not accurately distinguish between genotypes 1a and 1b in 5-10% of cases, because specific probes for both subtypes are lacking (2). Mistyping of genotype 6 as genotype 1 may influence the clinical management of patients, since genotype 6 viruses show a higher response to therapy than genotype 1 (4).

The new VERSANT HCV Genotype 2.0 assay is designed to increase the accuracy of HCV genotype/subtype identification by including core specific probes that discriminate between genotypes 1a, 1b and 6c-6l. Because of the retained phylogenetic information in this region, the core gene is one of the recommended regions for HCV genotyping (18). Although attempts have been made previously to design specific 5'UTR/core probes for genotype identification, these were not tested for genotype 6 viruses (25). In this study, we have tested the new VERSANT HCV Genotype 2.0 line probe assay with 73 samples obtained from SEA, of which 45% were of genotype 6 (33/73). A marked improvement in accurate genotyping is seen with the new assay as compared to the old assay that only targeted the 5'UTR (~96% versus 71%). As expected, the increase in precision is largely due to correct genotyping and subtyping of 6c-6l and 1a/b viruses, respectively. Our sample set not only contains genotype 6a, but also the new variants of genotypes 6d, 6e, 6f, 6i, 6l and 6n (Fig. 1). All of these were correctly identified by the VERSANT HCV Genotype 2.0 assay. With regard to the heterogeneity of genotype 6 sequences, 5'UTR/core line probe assay proved its ability to recognize a broad variety of genotype 6 viruses.

Only 4% (3/73) of the samples could not be genotyped. One of these showed a new pattern of genotype 6c-6l (Th52). Two other samples (Th6 and Th44) showed specific hybridization with probes of genotype 1b and 6c-6l. Although sequence variation between HCV strains offers an explanation as only few nucleotide changes could cause cross-reactivity with probes thought to be subtype-specific, this seems not very likely. Alternatively, these results could indicate the presence of genotypically different viruses in the samples. Attempts to confirm this by sequence analysis were not successful, possibly because the detection of mixed infections by sequence analysis of amplified fragments is not sensitive enough (6, 14). The hybridization patterns, indicating a

possible mixed infection of genotype 1 and 6 viruses, are not evident from the line patterns observed with the old version of the LiPA test.

One of the other commercial assays that is frequently used for genotyping HCV is the Trugene HCV 5'NC genotyping kit. In comparative studies between this assay and the previous VERSANT HCV Genotype test, the accuracy of genotyping was similar, although Trugene showed a slightly higher percentage of correct subtype identifications (5). The efficiency of Trugene in discriminating genotype 6 subtypes is currently not known. However, its exclusive use of the 5'UTR region for genotype identification points towards a reduced efficiency in the correct assignment of genotype 6c-6l viruses. Therefore, the new VERSANT HCV Genotype 2.0 assay, which also takes into account HCV core sequences, may actually be the most suitable current tool for routine HCV genotyping. This assay may not only be crucial for clinical evaluation of patients, but also for future epidemiological, evolutionary and pathogenesis studies.

ACKNOWLEDGEMENTS

The Commission of the European Community, HECSA Project (ICA4-CT-1999-10009), supported this work. We would like to thank Hilde van de Velde, An Cremmery, Dr. Vu Thuy Yen, Kavita Bedi and Apiradee Theamboonlers for their technical assistance.

REFERENCES

1. **Buoro, S., S. Pizzighella, R. Boschetto, L. Pellizzari, M. Cusan, R. Bonaguro, C. Mengoli, C. Caudai, M. Padula, P. Egisto Valensin, and G. Palu.** 1999. Typing of hepatitis C virus by a new method based on restriction fragment length polymorphism. *Intervirology* **42**:1-8.
2. **Chen, Z., and K. E. Weck.** 2002. Hepatitis C virus genotyping: interrogation of the 5' untranslated region cannot accurately distinguish genotypes 1a and 1b. *J. Clin. Microbiol.* **40**:3127-3134.
3. **Chinchai, T., J. Labout, S. Noppornpanth, A. Theamboonlers, B. L. Haagmans, A. D. Osterhaus, and Y. Poovorawan.** 2003. Comparative study of different methods to genotype hepatitis C virus type 6 variants. *J. Virol. Methods* **109**:195-201.
4. **Dev, A. T., R. McCaw, V. Sundararajan, S. Bowden, and W. Sievert.** 2002. Southeast Asian patients with chronic hepatitis C: the impact of novel genotypes and race on treatment outcome. *Hepatology* **36**:1259-1265.
5. **Halfon, P., P. Trimoulet, M. Bourliere, H. Khiri, V. de Ledinghen, P. Couzigou, J. M. Feryn, P. Alcaraz, C. Renou, H. J. Fleury, and D. Ouzan.** 2001. Hepatitis C virus genotyping based on 5' noncoding sequence analysis (Trugene). *J. Clin. Microbiol.* **39**:1771-1773.
6. **Hu, Y. W., E. Balaskas, M. Furione, P. H. Yen, G. Kessler, V. Scalia, L. Chui, and G. Sher.** 2000. Comparison and application of a novel genotyping method, semiautomated primer-specific and mispair extension analysis, and four other genotyping assays for detection of hepatitis C virus mixed-genotype infections. *J. Clin. Microbiol.* **38**:2807-2813.
7. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* **5**:150-163.
8. **Lee, J. H., W. K. Roth, and S. Zeuzem.** 1997. Evaluation and comparison of different hepatitis C virus genotyping and serotyping assays. *J. Hepatol.* **26**:1001-1009.
9. **McHutchison, J. G., S. C. Gordon, E. R. Schiff, M. L. Shiffman, W. M. Lee, V. K. Rustgi, Z. D. Goodman, M. H. Ling, S. Cort, and J. K. Albrecht.** 1998. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *N. Engl. J. Med.* **339**:1485-1492.
10. **Mellor, J., E. C. Holmes, L. M. Jarvis, P. L. Yap, and P. Simmonds.** 1995. Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J. Gen. Virol.* **76**:2493-2507.
11. **Mellor, J., E. A. Walsh, L. E. Prescott, L. M. Jarvis, F. Davidson, P. L. Yap, and P. Simmonds.** 1996. Survey of type 6 group variants of hepatitis C virus in Southeast Asia by using a core-based genotyping assay. *J. Clin. Microbiol.* **34**:417-423.
12. **Noppornpanth, S., T. X. Lien, Y. Poovorawan, S. L. Smits, A.D.M.E. Osterhaus, and B.L. Haagmans.** 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J. Virol.* **80**:7569-7577.

13. **Richter, S. S.** 2002. Laboratory assays for diagnosis and management of hepatitis C virus infection. *J. Clin. Microbiol.* **40**:4407-4412.
14. **Roque-Afonso, A. M., D. Ducoulombier, G. Di Liberto, R. Kara, M. Gigou, E. Dussaix, D. Samuel, and C. Feray.** 2005. Compartmentalization of hepatitis C virus genotypes between plasma and peripheral blood mononuclear cells. *J. Virol.* **79**:6349-6357.
15. **Salemi, M., and A. M. Vandamme.** 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J. Mol. Evol.* **54**:62-70.
16. **Sandres-Saune, K., P. Deny, C. Pasquier, V. Thibaut, G. Duverlie, and J. Izopet.** 2003. Determining hepatitis C genotype by analyzing the sequence of the NS5b region. *J. Virol. Methods* **109**:187-193.
17. **Silini, E., F. Bono, A. Cividini, A. Cerino, S. Bruno, S. Rossi, G. Belloni, B. Brugnetti, E. Civardi, L. Salvaneschi, and M.U. Mondelli.** 1995. Differential distribution of hepatitis C virus genotypes in patients with and without liver function abnormalities. *Hepatology* **21**:285-290.
18. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-973.
19. **Simmonds, P., J. Mellor, T. Sakuldamrongpanich, C. Nuchaprayoon, S. Tanprasert, E. C. Holmes, and D. B. Smith.** 1996. Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *J. Gen. Virol.* **77**:3013-3024.
20. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 1994. Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J. Gen. Virol.* **75**:1053-1061.
21. **Smith, D. B., J. Mellor, L. M. Jarvis, F. Davidson, J. Kolberg, M. Urdea, P. L. Yap, and P. Simmonds.** 1995. Variation of the hepatitis C virus 5' non-coding region: implications for secondary structure, virus detection and typing. The International HCV Collaborative Study Group. *J. Gen. Virol.* **76**:1749-1761.
22. **Smith, D. B., and P. Simmonds.** 1997. Review: molecular epidemiology of hepatitis C virus. *J. Gastroenterol. Hepatol.* **12**:522-527.
23. **Stuyver, L., R. Rossau, A. Wyseur, M. Duhamel, B. Vanderborght, H. Van Heuverswyn, and G. Maertens.** 1993. Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *J. Gen. Virol.* **74**:1093-1102.
24. **Stuyver, L., A. Wyseur, W. van Arnhem, F. Hernandez, and G. Maertens.** 1996. Second-generation line probe assay for hepatitis C virus genotyping. *J. Clin. Microbiol.* **34**:2259-2266.
25. **Stuyver, L., A. Wyseur, W. van Arnhem, F. Lunel, P. Laurent-Puig, J. M. Pawlotsky, B. Kleter, L. Bassit, J. Nkengasong, L. J. van Doorn, and G. Maertens.** 1995. Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. *Virus Res.* **38**:137-157.

26. **Theamboonlers, A., T. Chinchai, K. Bedi, P. Jantaramee, M. Sripontong, and Y. Poovorawan.** 2002. Molecular characterization of Hepatitis C virus (HCV) core region in HCV-infected Thai blood donors. *Acta Virol.* **46**:169-173.
27. **Tokita, H., H. Okamoto, H. Iizuka, J. Kishimoto, F. Tsuda, L. A. Lesmana, Y. Miyakawa, and M. Mayumi.** 1996. Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups. *J. Gen. Virol.* **77**:293-301.
28. **Tokita, H., H. Okamoto, F. Tsuda, P. Song, S. Nakata, T. Chosa, H. Iizuka, S. Mishiro, Y. Miyakawa, and M. Mayumi.** 1994. Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *Proc. Natl. Acad. Sci. USA* **91**:11022-11026.
29. **Weck, K.** 2005. Molecular methods of hepatitis C genotyping. *Expert Rev. Mol. Diagn.* **5**:507-520.
30. **Zein, N. N., J. Rakela, E. L. Krawitt, K. R. Reddy, T. Tominaga, and D. H. Persing.** 1996. Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. Collaborative Study Group. *Ann. Intern. Med.* **125**:634-639.
31. **Zylberberg, H., M. L. Chaix, and C. Brechot.** 2000. Infection with hepatitis C virus genotype 4 is associated with a poor response to interferon-alpha. *Ann. Intern. Med.* **132**:845-846.



CHAPTER 3

Identification of a Naturally Occurring Recombinant Genotype 2/6 Hepatitis C Virus

Suwanna Noppornpanth,¹ Truong Xuan Lien,² Yong Poovorawan,³ Saskia L. Smits,¹
Albert D. M. E. Osterhaus,¹ and Bart L. Haagmans¹

Department of Virology, Erasmus Medical Center, Rotterdam, The Netherlands,¹

*Department of Biological Analysis, Pasteur Institute Ho Chi Minh City, Ho Chi Minh
City, Vietnam²*

*Center of Excellence in Viral Hepatitis Research, Department of Pediatrics,
Chulalongkorn Hospital Bangkok, Thailand.³*

Journal of Virology, 80 (2006), p. 7569-77.

ABSTRACT

Hepatitis C viruses (HCVs) display a high level of sequence diversity and are currently classified into six genotypes and an increasing number of subtypes. Most likely, this heterogeneity is caused by genetic drift; evidence for recombination is scarce. To study the molecular heterogeneity of HCV in Vietnam, we analyzed 58 HCV RNA positive sera from Vietnamese blood donors by sequence analysis of the CORE and NS5B regions. Phylogenetic analyses revealed the presence of genotype 1 (38%), genotype 2 (10.3%) and genotype 6 viruses (51.7%). All samples showed concordant results except for two (D3 and D54). Sample D54 was a mixed infection of genotype 2i and 6h viruses. Whole genome analysis and bootscan analysis of sample D3 on the other hand, revealed a recombinant virus with genotype 2i and genotype 6p sequences at the 5' and 3' ends, respectively. The crossover point was located between nucleotide position 3405-3464 (numbering according to prototype strain HCV-H, M67463), at the NS2/NS3 junction. The identification of this naturally occurring recombinant virus strengthens the concept that recombination may play a role in HCV epidemiology and evolution. Furthermore, the location of the recombination breakpoint may be relevant for constructing infectious chimeric viruses.

INTRODUCTION

Hepatitis C virus (HCV), an important causative agent of acute and chronic hepatitis, is an enveloped plus-strand RNA virus that belongs to the family *Flaviviridae* (33). Its genome, approximately 9.4 kb in length, encodes both structural (Core, E1, E2) and nonstructural (p7, NS2, NS3, NS4a/b and NS5a/b) proteins in a single open reading frame (6). Short conserved untranslated regions (UTRs) located at the 5' and 3' ends of the genome are required for viral replication (9, 10). An internal ribosomal entry site in the 5'UTR is involved in protein translation (46).

Because of its heterogeneity, HCV is classified into six major genotypes and a large number of subtypes (36-38). Different genotypes display up to 30% sequence diversity, whereas subtypes vary more than 20% (37, 38). The variability is distributed unequally across the genome, with regions such as E1 and E2 displaying most sequence diversity, whereas 5'UTR and CORE sequences are more conserved (39). Several methods were developed to determine the HCV genotype such as serological genotyping, RT-PCR amplification with genotype specific primers, restriction fragment length polymorphism (RFLP) analysis, reverse hybridization assay and sequence analysis (28). Of these methods, sequence analysis of phylogenetically informative regions is more reliable for genotype and subtype identification. Genotyping of HCV is important for prediction of responses to, and in determining the duration of antiviral therapy (23, 28). Moreover, assessment of the distribution of genotypes in different parts of the world may help to understand the epidemiology and evolution of HCV.

The geographic distribution of HCV relates to different epidemic histories and routes of transmission. Some genotypes, such as 1a/b, 2 and 3a, are widely distributed to Western countries, United States and Japan as a result of transmission via blood transfusion and contaminated needles between intravenous drug users (IDUs) (2). These strains typically have limited sequence diversity, resulting from the recent introduction of a few strains from endemic areas (36). On the other hand, a more complex viral heterogeneity is observed in parts of Africa and South-East Asia. In Western Africa, HCV infection is caused predominantly by genotype 2, while genotype 1 and 4 are most prevalent in Central Africa (4, 24). In both areas, a notable divergence of subtypes, especially broad genotype 2 variety in Ghana, was reported (5). Similar observations

were made in Asia, where genotype 1, 3 and 6 dominate (33), and a large variety of genotype 6 was demonstrated in Vietnam, Thailand, Myanmar, China and Hong Kong (8, 25, 34). This large sequence divergence points to a long-term presence of HCV infection among local populations through variety of routes, including vertical, sexual and household contact transmission (35, 41) The sequence diversity of HCV is supposedly caused by a high mutation rate of the RNA-dependent RNA-polymerase during replication (41). Interestingly, recombination, commonly seen among RNA viruses, including other members of the family *Flaviviridae* (14, 20, 47, 51) is not thought to play a major role (39, 40, 49). The fact that recombinant forms of HCV have been observed in nature, such as the St. Petersburg strains containing sequences of both genotype 2k and 1b (18) and a possible 1a/1b recombinant virus in Peru (7), would suggest that all requirements for genetic exchanges could be met. However, only few chimeric HCV genomes were described, which may be due to detection limits of currently employed methods that are not suited for discovering recombinant strains. It is therefore possible that the true frequency of HCV recombination is underestimated (7, 18, 35), especially in endemic areas with a high prevalence of different HCV genotypes.

In the present study, we analyzed sera from 58 HCV RNA-positive blood donors from Ho Chi Minh City, Vietnam, in order to survey the molecular heterogeneity of HCV in this area. The CORE and NS5B genes of each sample were sequenced and phylogenetic analysis was performed. Two samples gave discordant results and were analyzed by full-length genome sequencing. Our results revealed one mixed genotype infection, whereas the other sample contained an intergenotypic recombinant virus.

MATERIALS AND METHODS

Serum samples. Fifty-eight previously identified HCV-positive sera from the Pasteur Institute Ho Chi Minh City were collected. All samples were obtained from blood donors between 2000-2002. Samples were kept at -80°C until further analysis.

Amplification of HCV CORE and NS5B regions. Viral RNA was extracted from 100 μl serum by the High Pure RNA isolation kit (Roche Diagnostics GmbH) and eluted with 30 μl water. cDNA was synthesized from 10 μl extracted RNA with Moloney murine leukemia virus reverse transcriptase (Invitrogen) and random hexamer primers, according to the instructions of the manufacturers. Core sequences (Fig.1A, nucleotides

339 to 695, according to the numbering system for reference strain HCV-H Genbank accession number M67463) were amplified by nested PCR with oligonucleotides s17, 410, 953, and 951 (Table 1), as described previously (24). NS5b sequences (Fig.1A, nucleotides 8283 to 8624) were amplified by nested PCR with primers Pr3, Pr4, Pr1, and Pr2 (Table 1) using conditions described in previous reports (32) with modifications. Briefly, in the first PCR, performed with Pr3 and Pr4, 1.5 mM MgCl₂ was used in the reaction mixture, and the following thermal profile: initial denaturation at 94°C for 90 s; 5 cycles of denaturation at 94°C for 30s, annealing at 64°C for 45 s and elongation at 72°C for 1 min, followed immediately by 30 cycles at 94°C for 30 s, 64°C with a drop of -0.5°C between each cycle and elongation at 72°C for 1 min. The last 5 cycles were performed at 94°C for 30 s, 48°C for 45 s and 72°C for 1 min. A final elongation at 72°C for 10 min was also included. The nested PCR, performed with primer Pr1 and Pr2, was carried out on 2 µl of first PCR product with the following thermal profile: 95°C for 7 min; 50 cycles at 95°C for 30 s, 63°C for 30 s and 72°C for 30 s, with a final elongation step at 72°C for 10 min. The amplified product were gel-purified and sequenced.

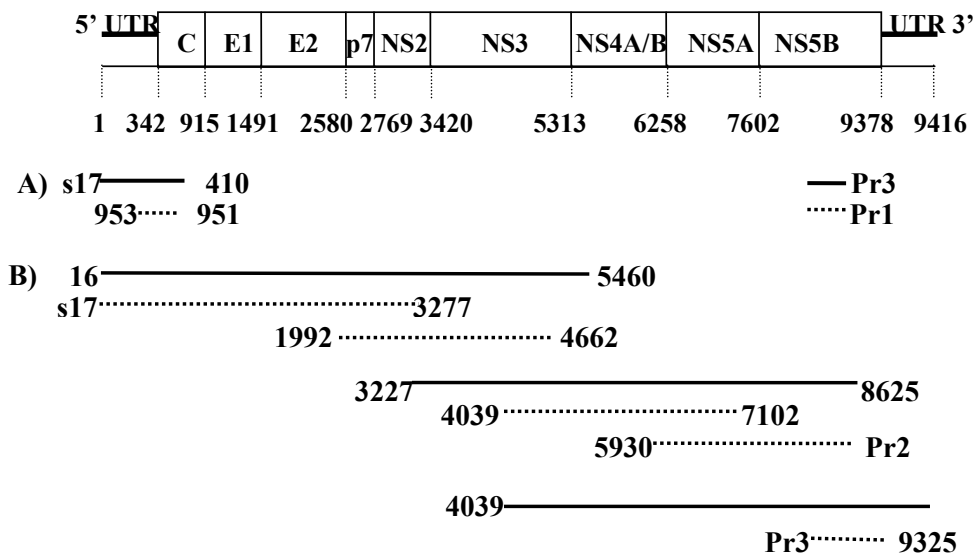


Figure 1 Hepatitis C virus genome organization and a schematic outline of the strategies used for RT-PCR amplification of (A) CORE and NS5B regions and (B) overlapping fragments covering the entire genome. Boxes represent the coding regions for the Core protein (C), envelope 1 and 2 (E1 and E2) proteins, p7, and nonstructural proteins 1-5 (NS1-NS5). Also indicated are the 5'- and 3' untranslated regions (5'UTR and 3'UTR). Black lines and dotted lines indicate the first- and second-round PCR products, respectively. Forward and reverse primers are presented at the 5' and 3' end of PCR products. The nucleotide positions and genetic organization are according to the numbering system for the prototype strain HCV-H (accession number M67463).

Table 1. Primers used for amplification and sequencing of HCV.

Primer	Sequence 5' – 3'	Position	Polarity	References
16 [†]	grggcgactccacat	16-33	Forward	this paper
s17 [†]	accatagatcactcccctgt	29-48	Forward	31
953	aggtctcgtagaccgtgcatcatg	321-344	Forward	24
951	caygtragggatcgatgac	697-716	Reverse	24
410	atgtaccccatgaggcggc	731-750	Reverse	24
693	aayttgggtaargtcacg	693-711	Forward	this paper
845	ccygggtgctcytytctat	849-868	Forward	this paper
1298	ctgggayatgatgatgaa	1298-1315	Forward	this paper
1992 [†]	ggntgyacvtggatgaa	1992-2008	Forward	this paper
2085	tsckgaarcartcvtrgg	2085-2104	Reverse	this paper
3277F [‡]	atggagaagaarvtyatcac	3277-3296	Forward	this paper
3277R [†]	ctratrabyttvdctccat	3277-3296	Reverse	this paper
3636	tggctcyacattggtrtacatytg	3636-3658	Reverse	this paper
4039 [¶]	ggcagyggbaaragyac	4039-4955	Forward	this paper
4662 [†]	crccygrtrtasccrgtcac	4662-4680	Reverse	this paper
4863	ccstengghatgttyga	4863-4879	Forward	this paper
A5310	cccaggtgctagtgcacgac	5304-5322	Reverse	31
5460 [†]	tagmrcayctccatctc	5460-5478	Reverse	this paper
5930 [‡]	agtsgeytyaartrcatg	5930-5948	Forward	this paper
6085	gctatbagyckgttcacca	6084-6103	Reverse	this paper
6800	tccebtgygarcebgag	6820-6836	Forward	this paper
7039	tggmggcargaratggg	7038-7054	Forward	this paper
7100 [‡]	aagmggvtcraargagt	7102-7118	Reverse	this paper
7600	tgvcgratkagrgagttgc	7681-7699	Reverse	this paper
8170	atgggrttctcctaygayac	8244-8263	Forward	this paper
Pr1	tggggatcccgtatgatacccgtgctttga	8245-8275	Forward	32
Pr3 [¶]	tatgayaccgctgytttgactc	8256-8278	Forward	32
Pr2 [‡]	ggcggaaattctggtcatagcctccgtgaa	8616-8645	Reverse	32
Pr4	gcngartayctvgtcatagcctc	8622-8644	Reverse	32
8625 [‡]	gartacctrgtcatagc	8625-8641	Reverse	this paper
9325 [¶]	ctacrgtaagtaggagtaggc	9325-9345	Reverse	this paper
3'UTR [¶]	aaaaaaaaaaaaaaaaaaaa	9418-9435	Reverse	this paper

[†] Primer used for amplification of cDNA fragment nt 17-5460

[‡] Primer used for amplification of cDNA fragment nt 3227-8625

[¶] Primer used for amplification of cDNA fragment nt 4039-9435

HCV full-length genome amplification. Total RNA was isolated from 100 μ l serum by proteinase K digestion and phenol-chloroform extraction as previously described (31) and dissolved in 20 μ l sterile water. To generate a long fragment of cDNA, reverse transcription was performed on 10 μ l RNA with Expand Reverse Transcriptase (Roche Diagnostics GmbH) and primers 5460, 8625 or 3'UTR (Table 1, Fig.1B), according to the manufacturer's instructions, in a total volume of 30 μ l at 42 °C for 2h. Ten- μ l volumes of cDNA were amplified with 2.5 units of Expand High Fidelity Enzyme mixture (Roche Diagnostics GmbH), 1x Expand PCR buffer, 0.4 mM of each dideoxynucleotide and 0.4 μ M outer primer pairs (Table 1) with a thermal profile that was described previously (31). First-round amplification from each cDNA reaction resulted in three overlapping PCR fragments (nucleotides 16 to 5460, 3227 to 8625 and 4039 to 9416, Fig.1B). Two- μ l volumes of amplified product were subjected to nested PCR with inner primers to generate overlapping genome fragments (nucleotides 17 to 3277, 1992 to 4662, 4039 to 7100, 5930 to 8645 and 8256 to 9325, Fig. 1B) under similar conditions as employed for the first PCR. Amplified products were gel-purified with the Qiaquick gel extraction kit (Qiagen) and ligated directly into plasmid pCR2.1 (TA cloning kit, Invitrogen). For each amplicon, five positive clones were selected and sequenced.

DNA sequencing and sequence analysis. Sequence reactions on PCR products and plasmids were performed using BigDye Terminator Cycle sequencing Ready Reaction kit (Applied Biosystems) and an ABI prism 3100 Autosequencer (Applied Biosystems). Sequencing primers were shown in Table 1. Multiple sequence alignments were generated with the BioEdit program (version 7.0.1). Phylogenetic analyses were conducted using the Kimura-2 parameter model and Neighbor-joining methods. Bootstrapping (1000 replicates) was tested and consensus trees were produced by program of MEGA version 3.0 (19).

Reference strains used in this study were obtained from Genbank: genotype 1a (M67463), 1b (D90208), 2a (AF169005, AB047644), 2b (AF238486), 2c (D50409), 2f (D49754), 2k (AB031663), 2i (L48499, X76411), 3a (D28917), 3b (D49374), 3k (10a, D63821), 4a (Y11604), 5a (Y13184), 6a (Y12083), 6b (D84262), 6c (7d, D37857, D37843), 6d (7b, D84263), 6e (7a, D87365, D88478), 6f (7c, D38078, D38079), 6g (11a, D63822), 6h (9a, D84265), 6i (9b, D37864, D37850), 6j (9c, D37862, D37848), 6k (8b,

D84264), 6l (8a, D87359, D88472), 6p (L38340, L38380). The sequence of hepatitis G-B virus (U22304) was used as an outgroup.

Similarity and bootscanning analyses of whole genome sequences were performed with the Simplot program (version 2.5 for Window 95/NT; available from the author, Stuart C. Ray, online (<http://sray.med.som.jhmi.edu/RaySoft/SimPlot/>)).

Nucleotide sequence accession numbers. New sequences reported in this study have been submitted to Genbank and have been assigned accession numbers DQ155444-DQ155501 (CORE sequences), DQ155502-DQ155559 (NS5B sequences), DQ155560 (D3 complete genome sequence), DQ155561 (D54 genotype 2i complete genome sequence), DQ155562 and DQ155566 (D54 5'UTR-to-CORE sequences).

RESULTS

Genotyping of HCV isolates from Ho Chi Minh City, Vietnam.

We obtained 58 serum samples from Vietnamese blood donors, which were previously found to be HCV antibody positive. Viral RNA served as template for RT-PCR amplification of the Core and NS5B sequences (Fig.1). To study HCV diversity, a phylogram, based on 333 nucleotides of the NS5B isolated from these 58 samples and 26 prototype strains, was constructed (Fig. 2A). This analysis revealed the presence of genotype 1 (38%), genotype 2 (10.3%) and genotype 6 (51.7%) viruses. All Vietnamese 1b isolates sub-clustered together with a bootstrap value of 80%, while most of the subtype 1a strains were grouped with the 1a reference sequence (Fig. 2A). Five samples were characterized as subtype 2a, and one strain (D54) was identified as subtype 2i. High levels of sequence divergence were observed among the genotype 6 viruses; 4 HCV's (D41, D53, D83, and D88) exhibited separate branches from prototype strains, supported by bootstrap values. These data are a showpiece example of why it is so difficult to subtype genotype 6 variants based on reference sequences.

To confirm the consistency of genotyping, we constructed a phylogram based on the 357 nucleotides of the CORE sequence (Fig. 2B). Although the HCV CORE sequences of Vietnamese genotype 1 and 2 strains showed less divergence compared to prototypes than in the NS5B sequence, the genotype 6 strains showed considerable variety. Overall, 56/58 samples showed concordant genotyping results. Interestingly, sample D3 was characterized as genotype 2i based on the CORE region but typed as a

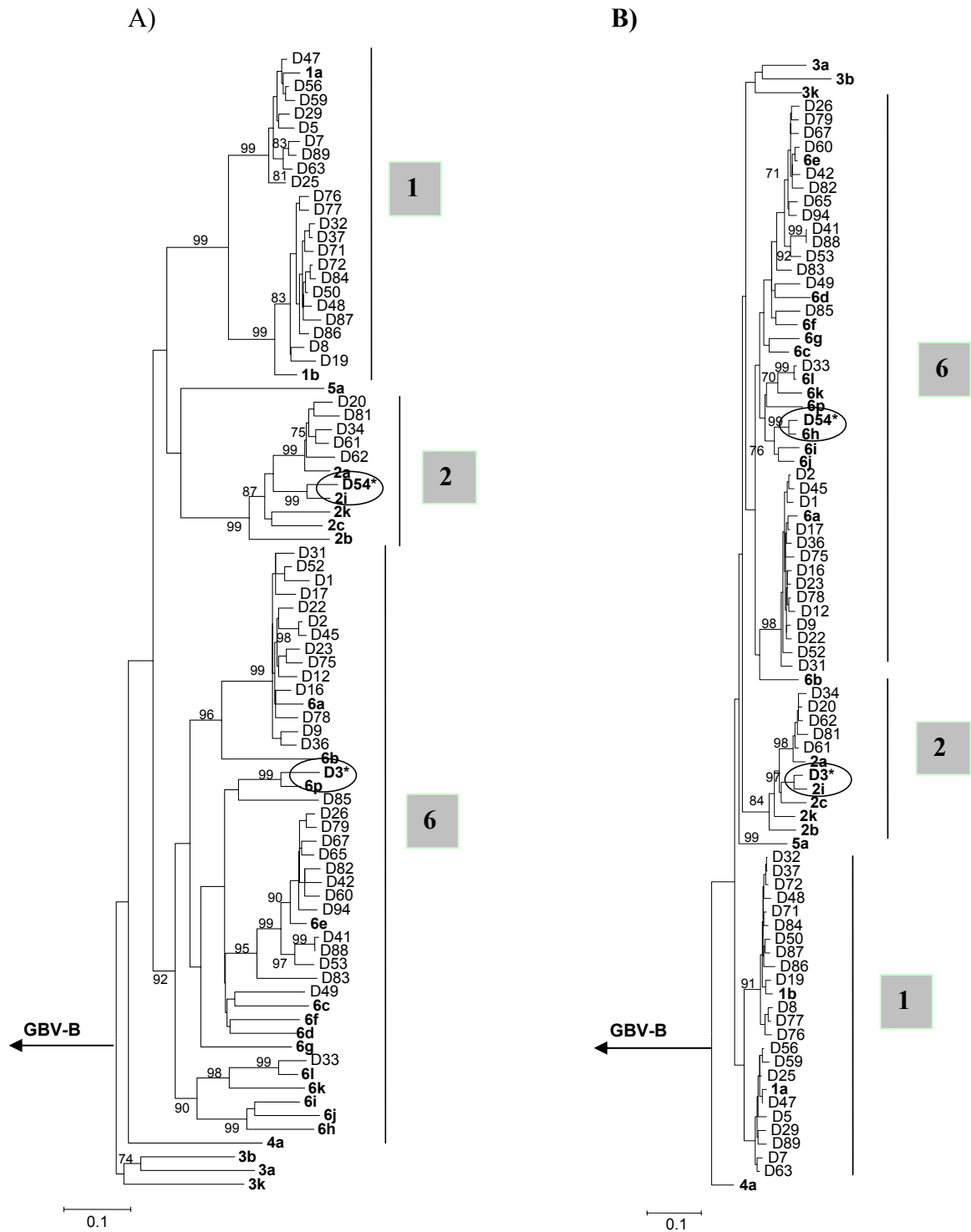


Figure 2 Rooted neighbor-joining trees depicting the phylogenetic relationships among HCV field variants and prototype strains. Trees were constructed for (A) a 333-nucleotide segment of the NS5B gene and (B) a 357 nucleotides of the Core gene, with the Kimura-2 parameter model. Confidence values (>70%) calculated by bootstrap analysis (1,000 replicates) are indicated at the major branching points. Branch lengths are drawn to scale. The prototype HCV strains obtained from Genbank are indicated by bold letters and Hepatitis G-B virus (U22304) was used as an outgroup sequence. Samples that gave discordant results in the phylogenetic trees are specified by asterisks and grouped to the related strain with circular mark.

genotype 6p after phylogenetic analysis of the NS5B sequence. Sample D54 on the other hand was classified as genotype 6h by analysis of the CORE sequence but grouped with genotype 2i after NS5B analysis.

Characterization of sample D54: mixed infection with HCV genotypes 2 and 6.

Conflicting phylogenetic data, such as obtained for the CORE and NS5B sequences in samples D3 and D54, are indicative for the presence of either a mixed infection with two different genotypes or an intergenotypic recombinant virus. Upon analysis with the 5'UTR reverse-hybridization line probe assay (INNO-LiPA HCV II), HCV from sample D3 was identified as genotype 2 and patient D54 was shown to have a mixed infection with HCVs of genotype 2 and genotype 1 or 6 variant (data not shown).

To further characterize sample D54, the 5'-UTR-to-CORE region (nucleotides 29-752) was amplified, cloned and sequenced. The alignment of CORE sequences obtained from sample D54 with prototype strains indicated that this patient had been double-infected with HCV genotype 2i (clone 20) and 6h (clone1) viruses (Fig.3). Complete genome sequences of both genotypes from sample D54 were determined following the strategy indicated in figure 1. No evidence for recombination was found. Phylogenetic analysis of CORE, E1 and NS5B sequences of HCV from sample D54 confirmed that one of the full-length sequences was of genotype 2i, whereas the other was of genotype 6h (data not shown).

Characterization of sample D3: HCV genotype 2/6 recombinant.

Sequence analysis of 10 individual clones of 5'UTR-CORE sequences from sample D3, confirmed the data from the INNO-LiPA HCV II line probe assay, suggesting the presence of a viral strain belonging to genotype 2i (data not shown). Because the data depicted in Fig. 2 strongly indicate the presence of both genotype 2i and 6p sequences, we suspected the presence of a recombinant virus. The full-length genome sequence of sample D3 was determined. In order to identify recombination breakpoints, we employed similarity plot and bootscanning analysis (Fig. 4). Since complete sequences of genotype 2i and genotype 6p were not found in sequence databases, we used genotype 6d (VN235, D84236) and the genotype 2i strain from sample D54 as reference sequences. Genotype 1a (HCV-H, M67463) was included as an outgroup sequence. Fig. 4 A shows the genetic distances between the reference strains and the

HCV genotype 2/6 recombinant

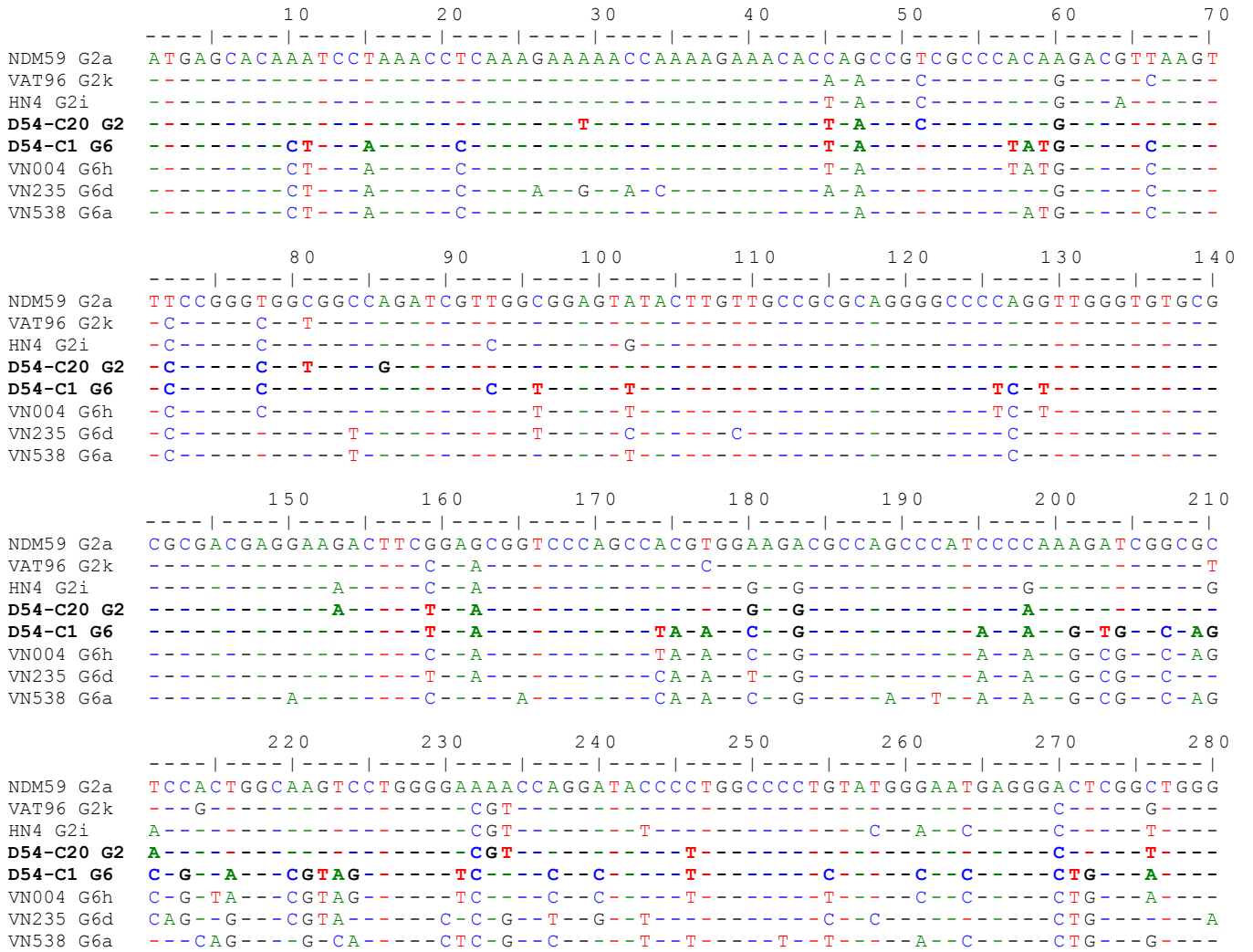
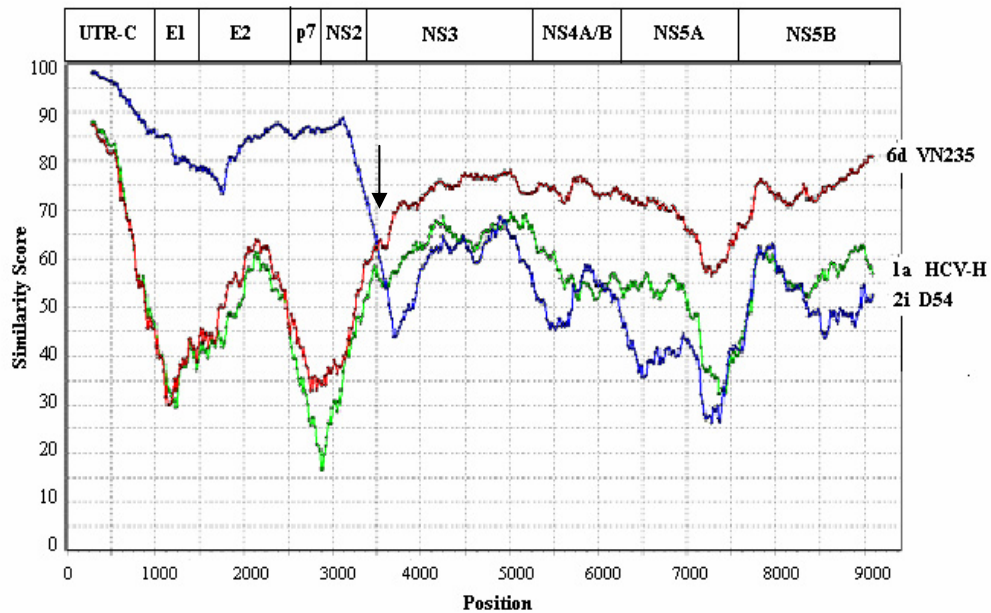


Fig. 3 Alignment of CORE sequences obtained from sample D54. Two representative clones, clone 1 (C1) and clone 20 (C20), were aligned with genotype 2a, 2k, 2i, 6a, 6d and 6h prototype strains. The D54 sequences are indicated in bold. Accession numbers of reference strains are indicated in Materials and Methods. Residues identical to the major sequence are indicated by dashed and numbering of the nucleotides starts at the first codon of CORE gene.

potential recombinant HCV strain of sample D3. The 5'UTR-to-NS2 region of the HCV strain of sample D3 is more related to subtype 2i sequences, whereas the NS3-to-NS5B region is more closely related to subtype 6d. Thus, the Simplot analysis (Fig. 4) corroborates the results from the phylogenetic trees (Fig. 2), providing evidence that the HCV strain in sample D3 is a recombinant virus. It is of note that the recombinant virus of sample D3 shows ~25% sequence divergence when compared with genotype 6d viruses in its C-terminal part of the genome.

A)



B)

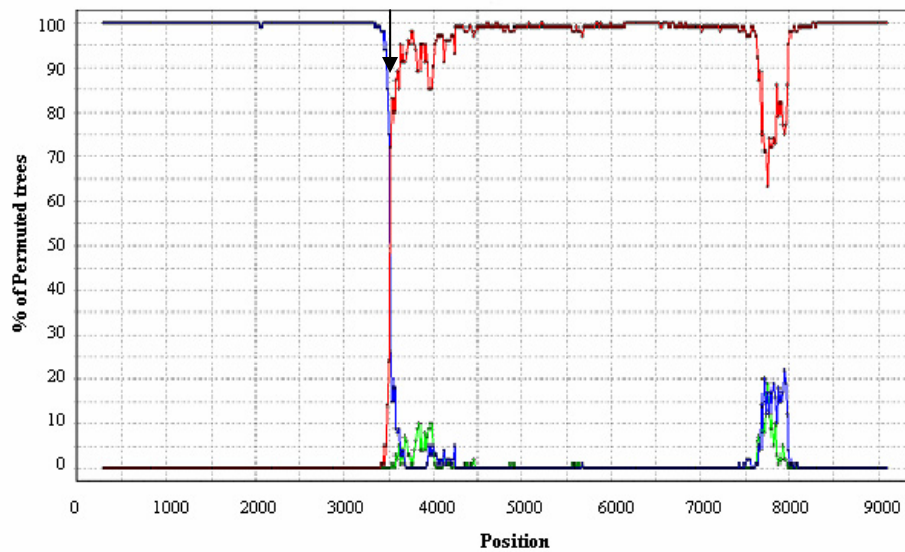


Fig. 4 (A) Similarity plots of the intergenotypic recombinant D3 strain and HCV genotype 1a (HCV-H), 2i (D54) and 6d (VN235) based on the entire genome using a window size of 600 bp, a step size of 20 bp and the maximum-likelihood parameter. Gaps were ignored and 1000 bootstrap replicates were used. (B) Bootscan plots showing the likelihood of clustering of the putative intergenotypic recombinant of sample D3 with reference strains, with 1000 bootstrap replicates and neighbour-joining tree analysis. As a reference, the genomic organization of HCV, drawn to scale, is depicted on top. Arrows indicate the recombination breakpoints. The nucleotide positions are numbered as described in the legend of Fig.1.

Bootscan analysis of the D3 HCV genome revealed the existence of only one recombination breakpoint, located within the NS2-NS3 region (Fig. 4B), between nucleotides 3405 - 3464 (Fig. 5A) and amino acids 1022 – 1042 (Fig. 5B). Interestingly, the recombination site is a conserved region at the amino acid level (Fig. 5B), which is cleaved by the NS2-NS3 protease (15, 43). Almost all cases of genetic exchanges in RNA viruses are thought to occur via copy-choice RNA recombination (26, 51). Template switching is supposedly aided by homology between the donor and acceptor strand and secondary RNA structures (1, 26). According to the proposed mechanism of HCV recombinant formation described by Kalinina et al. (17), the secondary RNA structure around the recombination site may enhance genetic exchanges *in vivo*. We could identify a stable hairpin structure (nucleotides 3423 to 3464) in the positive strand of genotype 6d strain VN235, that is not present in genotype 2i, which may have played a role in the creation of our recombinant genotype 2i/6 virus (data not shown).

To rule out the possibility of reverse transcription and PCR artifacts, we used a different reverse primer (4662, Table 1) to generate a cDNA fragment of HCV from sample D3 and amplified a short PCR product covering the NS2/NS3 region with primer 3277F and 4662 (position 3277 to 4662) (Fig.1). The results confirmed the recombination breakpoint.

DISCUSSION

Until recently, there was no evidence for recombination in hepatitis C viruses. Therefore, it has been implicitly assumed that HCV diversity was generated through genetic drift. Recently, however, both intergenotypic and intragenotypic recombinant HCVs have been identified (7, 18). In this paper, we have identified a second intergenotypic recombinant form of HCV.

Molecular heterogeneity of HCV in 58 HCV-positive serum samples from blood donors in Ho Chi Minh City were analysed. Phylogenetic analyses revealed the presence of genotype 1 (38%), genotype 2 (10.3%) and genotype 6 (51.7%) viruses. The Vietnamese genotype 1a and 1b viruses in our study, formed separate clusters and showed lower diversity compared to the genotype 6 strains. This may suggest that genotype 1 viruses were relatively recently introduced in Vietnam and were spread by

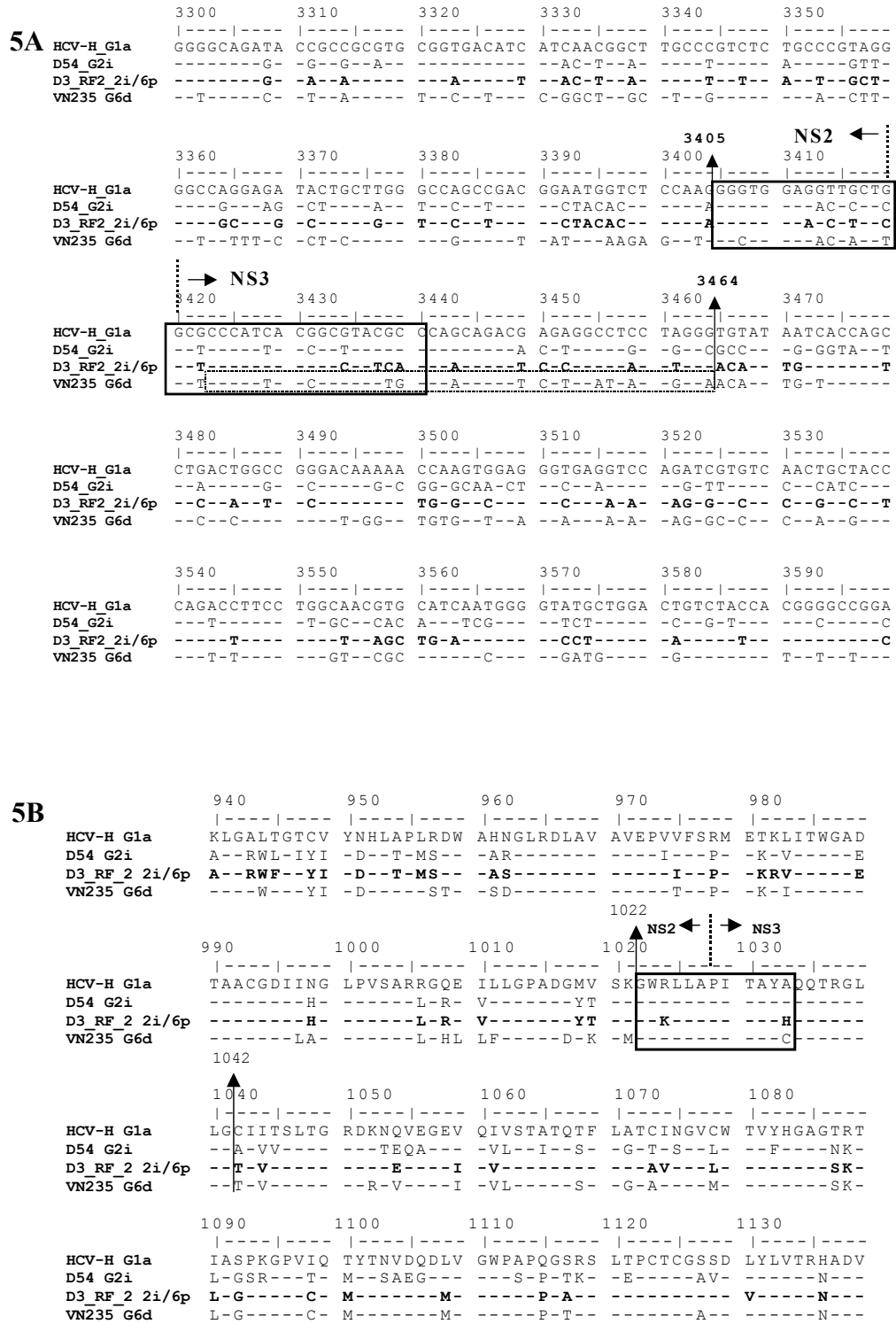


Fig. 5 Alignment of nucleotide (A) and amino acid (B) sequences of the NS2/NS3 region of the intergenotypic HCV recombinant of sample D3 and reference strains. Recombinant sequence is shown in bold. Boxes with solid lines indicate the cleavage site for the NS2/NS3 protease. The dotted box indicates the sequence forming the postulated hairpin proposed to promote recombination. The broken lines indicate the protease cleavage site of the NS2-NS3 protein. Vertical arrows overlap nucleotides 3405-3464, or amino acid 1022-1042, represented the recombination breakpoint. The nucleotide and amino acid positions are according to the numbering system described in the legend of Fig 1.

intravenous drug users and through use of contaminated blood products (35, 41). The higher heterogeneity of genotype 6 viruses would indicate, on the other hand, that these strains have circulated in Vietnam for a longer period of time (29, 41). The Vietnamese genotype 2a viruses clustered together and are related to subtype 2a from Japan. Interestingly, one of the strains that infected patient D54 was identified as a relatively rare subtype 2i, based on comparison with CORE, E1, and NS5B sequences of prototype 2i strain HN4. Subtype 2i was first identified in France in 1994 (30); the origin of the Vietnamese 2i virus is unknown. It is not unlikely that it was introduced in Vietnam by French citizens.

The most striking observation in our study is the identification of a HCV recombinant in Vietnam; the second intergenotypic recombinant that is described thus far. This virus contains sequences from subtype 2i and genotype 6. The genotype 6 region in this virus is probably most closely related to a new subtype 6p, of which recently a short NS5B sequence was reported (42). The new HCV recombinant thus can be classified as a candidate RF2_2i/6p strain. The recombination breakpoint is located between nucleotides 3405 and 3464. In the HCV polyprotein, this region is cleaved by the NS2-NS3 protease, and is reasonably conserved among different HCV genotypes (43). The copy-choice model, originally proposed in case of poliovirus recombination (20, 45), supports almost all possible mechanisms of RNA viruses described thus far. In this case, hybrid RNAs might be formed when the viral RNA-dependent RNA polymerase (RdRp) complex switches during replication. Our data suggest that both the high level of homology between donor and acceptor strand, and the presence of a hairpin structure in the region where the genetic exchange occurred, facilitated the generation of this recombinant.

It is worth noting that the recombination breakpoints in both identified intergenotypic HCV recombinants are located in the NS2/NS3 region of the genome (18). This may suggest that this region of the genome is especially suitable for creating viable intergenotypic recombinants *in vivo*. Moreover, this observation may aid researchers in designing chimeric HCV genomes for *in vitro* experiments, using the recently constructed infectious clone of HCV (21, 50, 54).

The frequency of successful recombination events is determined among others by (i) the properties inherent to the viral replicase, (ii) the odds of double infection of single

cells, and (iii) the viability of the recombinant progeny and the increase in fitness of the recombinant virus compared to its non-recombinant parents. Overall there is ample evidence that the generation of HCV recombinants is a rare event (35, 40, 49). On the one hand most recombinants generated may not be stable which is illustrated by the fact that chimeric HCV replicons, even constructed from closely related HCV subtypes 1a and 1b, often fail to replicate in cell culture (11, 13). Recent studies indicated that optimal HCV replication complexes composed of the various NS proteins, *cis*-acting RNA elements and cellular factors, are required for HCV replication (12, 22, 52). Possibly, recombinants that maintain successful replication complexes are scarce and they may not be able to compete with parental viruses. But even if new replication competent recombinant viruses are generated, they may be hampered in infectious virus particle formation as suggested by recent experiments with infectious full length genotype 1a/2a chimeras (21).

However, the fact that recombinant HCV viruses have been found (7, 18) would suggest that in principal the viral replicase is allowing the generation of chimeric genomes. The overlap in genotype distributions in many parts of the world would in theory increase the chance of double infections and thus the likelihood of recombination, especially in cases of frequent repeated needle-sharing over short time-intervals by infected drug users and the use of contaminated blood products. Although most of HCV recombinant strains might be selected out by natural selection (35, 49), the fact that recombinant viruses have been identified would suggest that at least some recombinants are viable.

Based on these findings we may currently underestimate the true frequency of HCV recombination. Recombination events in HCV are difficult to detect if they occurred between variants of the same subtype or between highly diverse subtypes. Even though there are some reports describing different genotype/subtype sequences of different regions obtained from single sample (27, 53), there are always doubts of mixed infection or contaminated samples (27). Moreover, most studies into HCV variability are based on analysis of single, short genomic regions making detection of potential recombinants unlikely. Therefore, the current methodology of sequencing short genomic regions as markers for entire viral genomes is challenged now, since it may not produce

accurate results of all genetic characteristics. Sequence analysis of whole HCV genomes would possibly increase the detection of natural HCV recombinants.

Recombination has been well documented for many RNA viruses (51), including other members of the family *Flaviviridae* (3, 16, 44, 47, 48, 51), of which evolution was for long supposed to be clonal, with diversity generated by the accumulation of mutations. The mere fact that recombinant HCV viruses have been identified (7, 18; this study), suggests that the assumption that HCV diversity is caused only via genetic drift is no longer tenable and that the role recombination plays in HCV evolution and biology demands serious consideration. Clearly, the fact that recombination occurs between subtypes and genotypes of HCV not only presents a serious impediment on vaccine development, but also on prediction of response to antiviral therapy.

ACKNOWLEDGEMENTS

The Commission of the European Community, HECSA Project (ICA4-CT-1999-10009), supported this work. We would like to thank you Dr. Vu Thuy Yen, Jan Brouwer and Paula van Luijt for their technical assistance and Dr. Byron Martina for valuable discussions.

REFERENCES

1. **Alejska, M., A. Kurzyniska-Kokorniak, M. Broda, R. Kierzek, and M. Figlerowicz.** 2001. How RNA viruses exchange their genetic material. *Acta. Biochim. Pol.* **48**:391-407.
2. **Alter, M. J.** 1999. Hepatitis C virus infection in the United States. *J. Hepatol.* **31** Suppl 1:88-91.
3. **Becher, P., M. Orlich, and H. J. Thiel.** 2001. RNA recombination between persisting pestivirus and a vaccine strain: generation of cytopathogenic virus and induction of lethal disease. *J. Virol.* **75**:6256-6264.
4. **Bukh, J., R. H. Purcell, and R. H. Miller.** 1994. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proc. Natl. Acad. Sci. USA* **91**:8239-8243.
5. **Candotti, D., J. Temple, F. Sarkodie, and J. P. Allain.** 2003. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J. Virol.* **77**:7914-7923.
6. **Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, R. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**:2451-2455.
7. **Colina, R., D. Casane, S. Vasquez, L. Garcia-Aguirre, A. Chunga, H. Romero, B. Khan, and J. Cristina.** 2004. Evidence of intratypic recombination in natural populations of hepatitis C virus. *J. Gen. Virol.* **85**:31-37.
8. **Doi, H., C. Apichartpiyakul, K. I. Ohba, M. Mizokami, and H. Hotta.** 1996. Hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand, and identification of novel subtypes of HCV major type 6. *J. Clin. Microbiol.* **34**:569-574.
9. **Friebe, P., and R. Bartenschlager.** 2002. Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *J. Virol.* **76**:5326-5338.
10. **Friebe, P., V. Lohmann, N. Krieger, and R. Bartenschlager.** 2001. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J. Virol.* **75**:12047-12057.
11. **Gates, A. T., R. T. Sarisky, and B. Gu.** 2004. Sequence requirements for the development of a chimeric HCV replicon system. *Virus. Res.* **100**:213-222.
12. **Gosert, R., D. Egger, V. Lohmann, R. Bartenschlager, H. E. Blum, K. Bienz, and D. Moradpour.** 2003. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J. Virol.* **77**:5487-5492.
13. **Gu, B., A. T. Gates, O. Isken, S. E. Behrens, and R. T. Sarisky.** 2003. Replication studies using genotype 1a subgenomic hepatitis C virus replicons. *J. Virol.* **77**:5352-5359.
14. **Guillot, S., V. Caro, N. Cuervo, E. Korotkova, M. Combiescu, A. Persu, A. Aubert-Combiescu, F. Delpeyroux, and R. Crainic.** 2000. Natural genetic exchanges between vaccine and wild poliovirus strains in humans. *J. Virol.* **74**:8434-8443.
15. **Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura, and K. Shimotohno.** 1993. Two distinct proteinase

- activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J. Virol.* **67**:4665-4675.
16. **Holmes, E. C., M. Worobey, and A. Rambaut.** 1999. Phylogenetic evidence for recombination in dengue virus. *Mol. Biol. Evol.* **16**:405-409.
 17. **Kalinina, O., H. Norder, and L. O. Magnius.** 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**:1853-1857.
 18. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius.** 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J. Virol.* **76**:4034-4043.
 19. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* **5**:150-163.
 20. **Lai, M. M.** 1992. RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**:61-79.
 21. **Lindenbach, B. D., M. J. Evans, A. J. Syder, B. Wolk, T. L. Tellinghuisen, C. C. Liu, T. Maruyama, R. O. Hynes, D. R. Burton, J. A. McKeating, and C. M. Rice.** 2005. Complete replication of hepatitis C virus in cell culture. *Science* **309**:623-626.
 22. **Lohmann, V., S. Hoffmann, U. Herian, F. Penin, and R. Bartenschlager.** 2003. Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J. Virol.* **77**:3007-3019.
 23. **McHutchison, J. G., S. C. Gordon, E. R. Schiff, M. L. Shiffman, W. M. Lee, V. K. Rustgi, Z. D. Goodman, M. H. Ling, S. Cort, and J. K. Albrecht.** 1998. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *N. Engl. J. Med.* **339**:1485-1492.
 24. **Mellor, J., E. C. Holmes, L. M. Jarvis, P. L. Yap, and P. Simmonds.** 1995. Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J. Gen. Virol.* **76 (Pt 10)**:2493-2507.
 25. **Mellor, J., E. A. Walsh, L. E. Prescott, L. M. Jarvis, F. Davidson, P. L. Yap, and P. Simmonds.** 1996. Survey of type 6 group variants of hepatitis C virus in Southeast Asia by using a core-based genotyping assay. *J. Clin. Microbiol.* **34**:417-423.
 26. **Nagy, P. D., and A. E. Simon.** 1997. New insights into the mechanisms of RNA recombination. *Virology* **235**:1-9.
 27. **Ndjomou, J., O. G. Pybus, and B. Matz.** 2003. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *J. Gen. Virol.* **84**:2333-2341.
 28. **Nolte, F. S.** 2001. Hepatitis C virus genotyping: clinical implications and methods. *Mol. Diagn.* **6**:265-277.
 29. **Pybus, O. G., M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey.** 2001. The epidemic behavior of the hepatitis C virus. *Science* **292**:2323-2325.
 30. **Qu, D., O. Hantz, M. Gouy, L. Vitvitski, J. S. Li, F. Berby, S. P. Tong, and C. Trepo.** 1994. Heterogeneity of hepatitis C virus genotypes in France. *J. Gen. Virol.* **75 (Pt 5)**:1063-1070.

31. **Rispeter, K., M. Lu, S. Lechner, A. Zibert, and M. Roggendorf.** 1997. Cloning and characterization of a complete open reading frame of the hepatitis C virus genome in only two cDNA fragments. *J. Gen. Virol.* **78 (Pt 11):**2751-2759.
32. **Sandres-Saune, K., P. Deny, C. Pasquier, V. Thibaut, G. Duverlie, and J. Izopet.** 2003. Determining hepatitis C genotype by analyzing the sequence of the NS5b region. *J. Virol. Methods* **109:**187-193.
33. **Shepard, C. W., L. Finelli, and M. J. Alter.** 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* **5:**558-567.
34. **Shinji, T., Y. Y. Kyaw, K. Gokan, Y. Tanaka, K. Ochi, N. Kusano, T. Mizushima, S. Fujioka, H. Shiraha, A. A. Lwin, Y. Shiratori, M. Mizokami, M. Khin, M. Miyahara, S. Okada, and N. Koide.** 2004. Analysis of HCV genotypes from blood donors shows three new HCV type 6 subgroups exist in Myanmar. *Acta Med. Okayama* **58:**135-142.
35. **Simmonds, P.** 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. *J. Gen. Virol.* **85:**3173-3188.
36. **Simmonds, P.** 1999. Viral heterogeneity of the hepatitis C virus. *J. Hepatol.* **31 Suppl 1:**54-60.
37. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42:**962-973.
38. **Simmonds, P., E. C. Holmes, T. A. Cha, S. W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea.** 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):**2391-2399.
39. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 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):**1053-1061.
40. **Smith, D. B., and P. Simmonds.** 1997. Review: molecular epidemiology of hepatitis C virus. *J. Gastroenterol. Hepatol.* **12:**522-527.
41. **Stumpf, M. P., and O. G. Pybus.** 2002. Genetic diversity and models of viral evolution for the hepatitis C virus. *FEMS Microbiol. Lett.* **214:**143-152.
42. **Stuyver, L., A. Wyseur, W. van Arnhem, F. Lunel, P. Laurent-Puig, J. M. Pawlotsky, B. Kleter, L. Bassit, J. Nkengasong, L. J. van Doorn, and G. Maertens.** 1995. Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. *Virus Res.* **38:**137-157.
43. **Suzuki, R., T. Suzuki, K. Ishii, Y. Matsuura, and T. Miyamura.** 1999. Processing and functions of Hepatitis C virus proteins. *Intervirology* **42:**145-152.
44. **Tolou, H. J., P. Couissinier-Paris, J. P. Durand, V. Mercier, J. J. de Pina, P. de Micco, F. Billoir, R. N. Charrel, and X. de Lamballerie.** 2001. Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. *J. Gen. Virol.* **82:**1283-1290.
45. **Tolskaya, E. A., L. I. Romanova, V. M. Blinov, E. G. Viktorova, A. N. Sinyakov, M. S. Kolesnikova, and V. I. Agol.** 1987. Studies on the recombination between RNA genomes of poliovirus: the primary structure and

- nonrandom distribution of crossover regions in the genomes of intertypic poliovirus recombinants. *Virology* **161**:54-61.
46. **Tsukiyama-Kohara, K., N. Izuka, M. Kohara, and A. Nomoto.** 1992. Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.* **66**:1476-1483.
 47. **Twiddy, S. S., and E. C. Holmes.** 2003. The extent of homologous recombination in members of the genus *Flavivirus*. *J. Gen. Virol.* **84**:429-440.
 48. **Uzcategui, N. Y., D. Camacho, G. Comach, R. Cuello de Uzcategui, E. C. Holmes, and E. A. Gould.** 2001. Molecular epidemiology of dengue type 2 virus in Venezuela: evidence for in situ virus evolution and recombination. *J. Gen. Virol.* **82**:2945-2953.
 49. **Viazov, S., A. Widell, and E. Nordenfelt.** 2000. Mixed infection with two types of hepatitis C virus is probably a rare event. *Infection* **28**:21-25.
 50. **Wakita, T., T. Pietschmann, T. Kato, T. Date, M. Miyamoto, Z. Zhao, K. Murthy, A. Habermann, H. G. Krausslich, M. Mizokami, R. Bartenschlager, and T. J. Liang.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* **11**:791-796.
 51. **Worobey, M., and E. C. Holmes.** 1999. Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **80 (Pt 10)**:2535-2543.
 52. **You, S., D. D. Stump, A. D. Branch, and C. M. Rice.** 2004. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J. Virol.* **78**:1352-1366.
 53. **Yun, Z., C. Lara, B. Johansson, I. Lorenzana de Rivera, and A. Sonnerborg.** 1996. Discrepancy of hepatitis C virus genotypes as determined by phylogenetic analysis of partial NS5 and core sequences. *J. Med. Virol.* **49**:155-160.
 54. **Zhong, J., P. Gastaminza, G. Cheng, S. Kapadia, T. Kato, D. R. Burton, S. F. Wieland, S. L. Uprichard, T. Wakita, and F. V. Chisari.** 2005. Robust hepatitis C virus infection in vitro. *Proc. Natl. Acad. Sci. USA* **102**:9294-9299.



CHAPTER 4

Characterization of Hepatitis C Virus Deletion Mutants Circulating in Chronically Infected Patients

Suwanna Noppornpanth,¹ Saskia L. Smits,¹ Truong Xuan Lien,² Yong Poovorawan,³
Albert D. M. E. Osterhaus,¹ and Bart L. Haagmans¹

Department of Virology, Erasmus Medical Center, Dr. Molewaterplein 50, 3015 GE, Rotterdam, The Netherlands¹

Department of Biological Analysis, Pasteur Institute Ho Chi Minh City, 167 Pasteur, Ho Chi Minh City, Vietnam²

Center of Excellence in Viral Hepatitis Research, Department of Pediatrics, 10330 Rama IV, Chulalongkorn Hospital Bangkok, Thailand³

Journal of Virology, 81(2007), p. 12496-503

ABSTRACT

Hepatitis C virus (HCV) has a linear positive-stranded RNA genome of ~9600 nucleotides in length and displays a high level of sequence diversity caused by high mutation rates and recombination. However, when we performed long distance reverse transcription PCRs on HCV RNA isolated from serum of chronic HCV patients not only full-length HCV genomes but also HCV RNAs which varied in size from 7600 to 8346 nucleotides and contained large in frame deletions between E1 and NS2 were amplified. Carefully designed control experiments indicated that these deletion mutants are a bona fide natural RNA species, most likely packaged in virions. Moreover, deletion mutants were detected in sera of patients infected with different HCV genotypes. We observed that 7/37 (18.9%) of genotype 1, 5/43 (11.6%) of genotype 3 and 4/13 (30.7 %) of genotype 6 samples contained HCV deletion mutant genomes. These observations further exemplify HCV's huge genetic diversity and warrant studies to explore their biological relevance.

INTRODUCTION

Hepatitis C virus (HCV) is an enveloped virus that belongs to the Family *Flaviviridae*. Its linear positive-stranded RNA genome of approximately 9.6 kb in length encodes both structural (Core, E1, E2) and nonstructural (p7, N2, NS3, NS4a/b and NS5a/b) proteins in a single open reading frame, flanked by short conserved untranslated regions (UTRs) located at the 5' and 3' ends of the genome required for viral replication and protein translation (5, 7, 8). One of the most striking characteristics of HCV is its capacity to cause a chronic liver disease in a high percentage of individuals (24). To do so, HCV encodes proteins which promote persistence and sequence variation, especially in the envelope genes (E1 and E2), results in escape from adaptive immune responses (9, 27).

Lack of proof reading ability of the viral RNA-dependent RNA polymerase is the driving force behind HCV's genetic diversity. As a result of the large amounts (10^{12}) of virions produced each day in chronic hepatitis C patients and the rate of incorrect nucleotide insertions which reaches the order of 10^{-3} to 10^{-4} base substitutions per site per year, HCV quasispecies are generated (22, 26). Recombination may be one other mechanism by which genetic diversity is driven, given the recent identification of natural occurring intergenotypic recombinant viruses (13, 21). Because of the huge genetic diversity, HCVs are currently categorized into six major genotypes and more than 80 subtypes (25).

HCV genetic variation has been studied in relation to epidemiology, response to antiviral therapy and clinical parameters, using different techniques that have focused on short genomic regions. However, analysis of full-length viral genomes may be necessary to better understand the characteristics of HCV. Previously, we analyzed sera from HCV RNA-positive blood donors from Ho Chi Minh City, Vietnam in order to analyze the molecular heterogeneity of HCV in South East Asia (21). Based on sequence analysis of Core and NS5b regions in a set of sera, two samples were identified which contained viruses of different genotypes. Whole genome analysis and bootscan analysis of one particular sample revealed a recombinant virus with genotype 2i and genotype 6p sequences at the 5' and 3' ends, respectively. Partial characterization of the other sample revealed the presence of a full-length genotype 2i virus and another genotype 6 virus,

which could be characterized only partially. In the present study, we further characterized this virus by a large fragment amplification method. Surprisingly, we identified a naturally occurring HCV deletion mutant genotype 6h virus that contained a large in-frame deletion of E1 and E2 genes. Further studies confirmed the existence of circulating HCV E1-E2 deletion mutants in a substantial percentage of chronic hepatitis C patients.

MATERIALS AND METHODS

Samples. HCV-positive sera of blood donors from Ho Chi Minh City, Vietnam and Bangkok, Thailand were obtained during the period from 2000 to 2002. Samples were genotyped based on sequence analysis as described previously (21). Thirteen samples of HCV genotype 6 were selected and characterized in this study. Other sera from chronic hepatitis C were obtained from the Erasmus MC, Rotterdam; thirty-seven plasmas of HCV genotype 1 and forty-three samples of genotype 3 were randomly selected for this study. All samples were kept at -80°C until further analysis.

RNA extraction and large fragment PCR amplification. Total RNA was isolated from 100 μl serum by proteinase K digestion and phenol-chloroform extraction as previously described (23) and dissolved in 20 μl sterile water. To generate a long fragment of cDNA, reverse transcription (RT) was performed on 10 μl RNA with Expand Reverse Transcriptase (Roche Diagnostics GmbH) and primers 5460, 8625 or 3'UTR, according to the manufacturer's instructions, in a total volume of 30 μl at 42°C for 2h as described previously (21). Ten- μl volumes of cDNA were amplified with 2.5 units of Expand High Fidelity Enzyme mixture (Roche Diagnostics GmbH), 1x Expand PCR buffer, 0.4 mM of each dideoxynucleotide and 0.4 μM outer primer pairs with a thermal profile that was described previously (21). First-round amplification from each cDNA reaction resulted in three overlapping PCR fragments (primers 16 and 5460, 3227F and 8625, 4039 and 3'UTR). Next, 2- μl volumes of amplified product were subjected to nested PCR with inner primers to generate overlapping genome fragments (primers s17 and 3277R, 66 (5'-tcccgagagcctatag-3') and 3636, 1992 and 4662, 4039 and 7100, 5930 and Pr2, Pr3 and 9325) under similar conditions as employed for the first PCR. Amplified products were gel-purified with the Qiaquick gel extraction kit (Qiagen)

and ligated directly into plasmid pCR2.1 (TA cloning kit, Invitrogen). For each amplicon, five positive clones were selected and sequenced.

To determine the sensitivity of HCV deletion mutants to RNase, we performed an assay as previously described (14). HCV plasma samples were treated with or without 0.1% Triton X-100 in PBS at 37°C for 1 hr, subsequently 5 Units of RNase A (Qiagen) was added and incubation continued for 1 hr. In order to degrade RNase A before RNA extraction, Proteinase K was added to each reaction and incubated at 37 °C for 15 min prior to adding lysis buffer. RNA was extracted as previously described (23) and was analyzed by RT-PCR.

Large fragment PCR and deletion mutant amplification using specific primers. In order to avoid mismatch of degeneracy primers, specific primers were designed for large fragment PCR of specific samples. Moreover, primers overlapping the deletion region of specific samples were designed and used for deletion mutant amplification described in this study (Table 1). Conditions for RT-PCR were identical to the protocol described above.

TABLE 1 Primers used for amplification HCV of specific samples

Primer ^a	Sequence (5' to 3')	Position ^b	Polarity	Sample
3277R	CTRATRABYTTVDTCTCCAT	3277-3296	Reverse	All
3277RP21	GTGATGATCTTGGTCTCCAT	3277-3296	Reverse	P21
3277RR30	GTGATGACCTTAATTTCCAT	3277-3296	Reverse	R30
3277RD33	GTCATGAGCTTGGTCTCCAT	3277-3296	Reverse	D33
3277RD88	GTGATAACCTTCTTCTCCAT	3277-3296	Reverse	D88
DelFP21	TGCGGATCTGTTTTCTGTG	1182//2538	Forward	P21
DelFR30	GCTGATGACGTTATGATGG	987//2253	Forward	R30
DelFD33	CTATGAGGCGGATATACC	980//2389	Forward	D33
DelFD88	CCCTGCACTATCCCCAAC	913//2757	Forward	D88
DelFD54 [†]	GCAGCATCGGTTATGAGG	970//3003	Forward	D54

^{a†}, Specific primer used for amplification of sample D54 genotype 6h.

^{b//}, Nucleotide position of specific primer designed over the cross-junction site of deletion region. The nucleotide position is according to the numbering system for the prototype strain HCV-H (accession number M67463).

cDNA synthesis using high thermal Reverse Transcriptase. To prevent HCV RNA secondary structure formation during RT, we performed cDNA synthesis using high thermostability reverse transcriptase enzyme. Transcriptor Reverse Transcriptase (Roche) and SuperscriptTM III Reverse Transcriptase (Invitrogen) were used for cDNA synthesis, according to the manufacturer's instructions. Protector RNase Inhibitor (Roche) was used in this experiment because of its heat resistant property. Conditions of RT-PCR are identical to the protocol described above, except for the incubation times; transcriptor RT was incubated at 50°C for 5 min and 55°C for 50 min while Superscript III was used at 55°C for 50 min.

Southern blot hybridization. Ten µl of PCR amplification products were loaded on 0.8% agarose-Tris-boric-EDTA (TBE) buffer gel. DNA samples were denatured in 0.5 N NaOH and 1.5 M NaCl for 45 min, neutralized in 1M Tris-HCl (pH 7.4) and 1.5 M NaCl for 45 min before transfer to Hybond N+ membranes (Amersham) by electroblotting (Amersham) in 1x TBE buffer according to the manufacturer. The blots were air dried for 30 min and UV-irradiated for 10 min. Hybridization procedure was conducted as described (6) with biotin-labeled probe specific for 5'UTR (5'-attccgggtgactcaccgggtccg-3', nucleotides 149-174) at 50 °C for 3 hrs. Membranes were washed and incubated with Streptavidin-POD conjugated (Roche) at 42°C for 45 min. After several washes the blots were visualized with enhanced chemiluminescence detection reagents (Roche) according to the protocol of the manufacturer, and by exposure to hyperfilm (Amersham) for 1 to 5 min.

Preparation of synthetic complete and deletion mutant HCV RNA. HCV large fragment amplicons were cloned into plasmid pCR 2.1 (TA cloning kit, Invitrogen) under control of the T7 promoter. Plasmids were linearized with BamHI and purified using Qiaquick PCR purification (Qiagen) and *in vitro* transcription was performed as described elsewhere (2). Reaction mixtures contained ~ 1µg of linearized DNA template, 5.0 mM rNTP-mix, 40 U RNaseOut (Invitrogen), 1x reaction buffer, and 50 U T7 RNA polymerase (Invitrogen) in a total volume of 20 µl. After 2 hrs at 37 °C, 5 U DNase I (Qiagen) was added and incubation prolonged for 1 hr. DNase-treated synthetic RNA was purified by phenol-chloroform extraction and used as a template for RT-PCR reaction.

DNA sequencing and sequence analysis. Sequence reactions on PCR products and plasmids were performed using BigDye Terminator Cycle sequencing Ready Reaction kit (Applied Biosystems) and an ABI prism 3100 Autosequencer (Applied Biosystems). Multiple sequence alignments were generated with the BioEdit program (version 7.0.1).

Nucleotide sequence accession numbers. New sequences reported in this study have been submitted to Genbank and have been assigned accession numbers: RP21.complete (EF420126), RP21.deletion (EF420127), RP30.complete (EF420128), RP30.deletion (EF420129), D88.complete (EF420130), D88.deletion (EF420131), D33.complete (EF420132), D33.deletion (EF420133) and D54 (EF420134).

RESULTS

Identification of a HCV genotype 6 deletion mutant. In order to study the genetic diversity of HCV in South East Asia, we analyzed serum samples obtained from Vietnamese blood donors. Previously, we reported that cloning of 5'UTR-to-Core region of one specific sample (D54) revealed the presence of genotype 2i and 6h strains and the whole genome of the genotype 2i virus was published previously (21). However, short fragment amplification of nucleotides 17 to 2085 (according to numbering system of prototype strain, M62321, genotype 1b) and 1992 to 3277 only amplified genotype 2i sequences; genotype 6h sequences were never obtained (21). Remarkably, large fragment amplification of nucleotides 17 to 3277 always amplified a smaller product of approximately 1200 bps. A subsequent effort to amplify a fragment from nucleotides 66 to 3636 again revealed the amplification of a small PCR product (approximately 1500 bps). In contrast, some other samples such as D42 showed the expected amplicon of 3600 bps (Fig. 1A). Southern blot hybridization using a HCV specific probe, revealed the specificity of the PCR products. Subsequent nucleotide sequence analysis of the obtained amplicons revealed a 2022 nucleotide (674 amino acid) in-frame deletion in E1/E2/p7 and part of the NS2 region (Fig. 1B) in the genotype 6h strain. Because of the nature of the PCR-method, the smaller genotype 6h fragment is amplified preferentially over the larger strain 2i fragment. Thus, both a full-length 2i strain and a deletion mutant of the 6h strain seem to be present in sample D54.

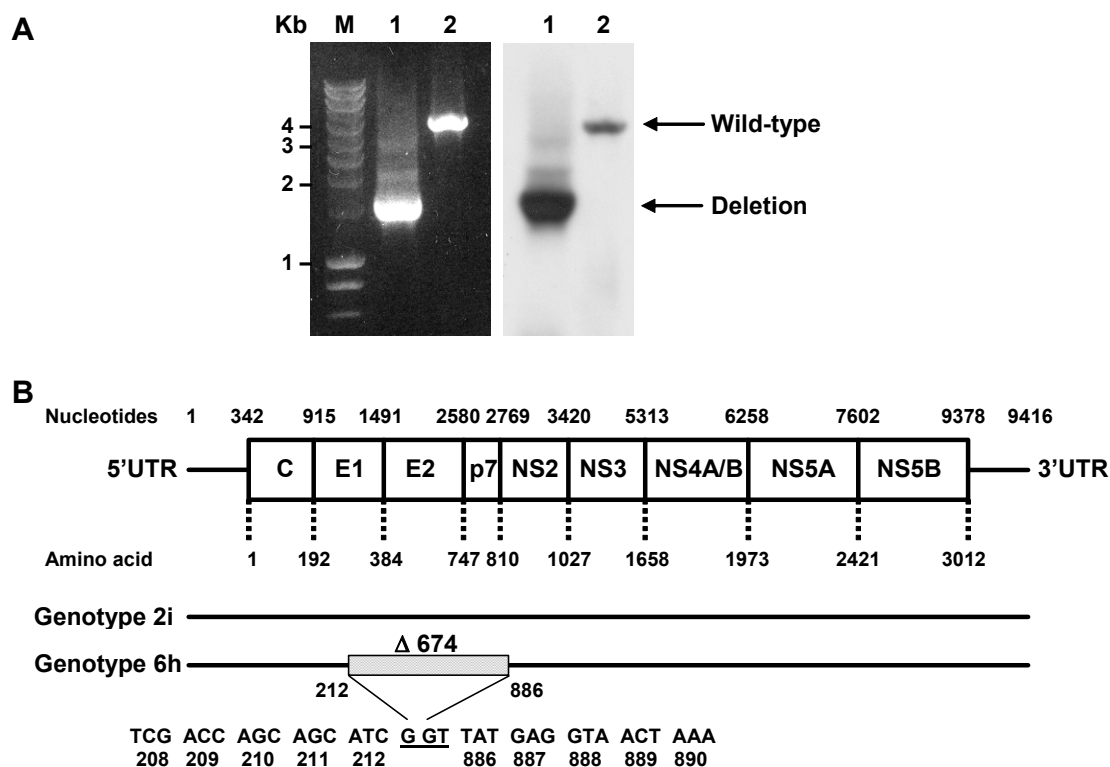


Figure 1. Identification of a HCV genotype 6h virus with a deletion in the E1-NS2 region in a patient (D54) double infected with genotypes 2i and 6h. (a) RT-PCR analysis (nucleotides 66-3636) of HCV RNA from sample D54 (lane 1) and sample D42 (lane 2) and subsequent southern blot analysis are shown in the left and right panel, respectively. Amplicons were separated in 0.8% agarose gels along with a marker (M) for which the molecular weights are given on the left side in kilobasepairs (Kb). The deletion mutant and wild-type HCV genome are indicated on the right side. (b) The upper panel depicts the genome organization of HCV with boxes indicating the coding regions for the Core protein (C), envelope 1 and 2 (E1 and E2) proteins, p7, and nonstructural proteins 1 to 5 (NS1 to NS5). Also indicated are the 5'- and 3'untranslated regions (5'UTR and 3'UTR) and the nucleotide and amino acid numbers, according to the numbering system for the prototype strain HCV-H (accession number M67463). In the lower panel, black lines indicate the D54 full-length (genotype 2i) and deletion (genotype 6h) sequences, respectively. The shaded box displays the position of the in-frame deletion region ($\Delta 674$).

Detection of HCV deletion mutants in chronically infected hepatitis C patients. To determine whether HCV deletion genomes can be detected in other hepatitis C patients, we randomly selected plasma samples obtained from HCV genotype 1, 3 and 6 infected blood donors or chronic hepatitis C patients and performed a long fragment amplification of the HCV genome (nucleotides 66 to 3636). We observed that 7/37 (18.9%) of genotype 1, 5/43 (11.6%) of genotype 3 and 4/13 (30.7 %) of genotype 6 samples

contained HCV deletion mutant genomes. RT-PCR analysis performed on at least two independent RNA isolations from the same patient sample resulted in the same deletion products. Four samples (genotype 1, 3 and genotype 6) were selected and found to contain deletion mutants of varying size (Fig. 2A). In contrast, amplicons generated by primer 1992 (located in the deleted region) and 3636, showed bands of the expected size (1644 bps) for all samples, indicating that non-deleted genomes were present in plasma as well (Fig. 2B). HCV specificity of the amplified bands was confirmed by southern blot analysis (data not shown). Sequencing of the fragments confirmed the presence of in-frame deletions in the E1-p7 region (Fig. 2C). Sequence analysis of 3-5 clones per sample revealed that the sequence of the deletion junction was identical in all clones for a given sample. Moreover, there were minor differences in the sequence of regions flanking the deletion junctions, implying independent quasispecies diversification of the deletion variants relative to their wild-type parent (Fig.3).

Experimental confirmation of HCV E1-E2 deletion mutants. To confirm that the deletion genomes were not RT-PCR artifacts, we performed a wide array of experiments to prove that the deletion mutant genomes are bona fide RNA species present in the serum of chronically infected patients. First of all, the initial large fragment PCR analysis utilized degenerate primers for amplification (21). To avoid the possibility of primer mismatch during amplification, reverse primers 3277R (nucleotides 3277-3296) specific for sample P21 (genotype 1), R30 (genotype 3), D33, and D88 (genotype 6) were used (Table 1). Subsequent analysis using these specific primers revealed similar results as observed in the previous experiment (Fig. 4A). In some but not all experiments the full-length genome could be detected, but at a lower level given the preferential amplification of the deletion mutant genomes. Moreover, amplification of the 5'UTR-to-Core (nucleotides 17 to 750) and NS3-to-NS5 (nucleotides 4039-8625) regions showed no evidence for the presence of deletions in these parts of the HCV genome (Fig. 4B-C). Next, we designed isolate specific primers overlapping the deleted region (Table 1). As presented in Fig. 4D, these primers amplified HCV specific products with appropriate lengths in the samples they were designed for, but not in others (not shown).

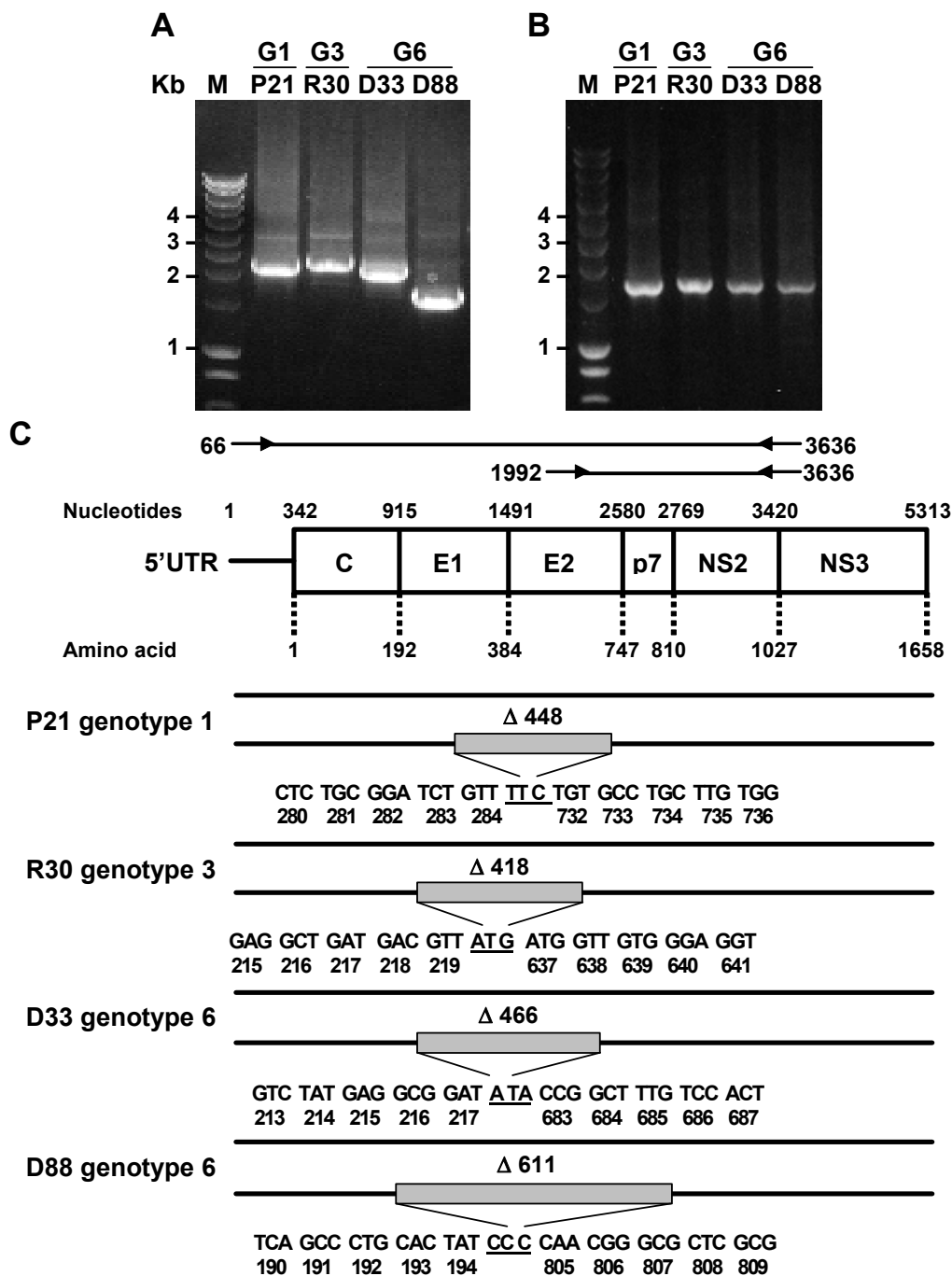


Figure 2. HCV deletion mutants occur in the serum of chronically HCV-infected genotype 1, 3, and 6 patients. RT-PCR products obtained from HCV genotype 1 (G1; P21), genotype 3 (G3; R30) and genotype 6 (G6; D33, D88) serum samples, using degenerate primers for amplification of HCV nucleotides 66 to 3636 (a) and 1992 to 3636 (b), were separated in 0.8% agarose gels. The molecular weight marker (M) is indicated on the left side in kilobasepairs (Kb). (c) A schematic representation of complete and deletion genomes determined for the different HCV genotypes. The upper panel depicts the genome organization of HCV as in Fig. 1. Forward and reverse primers to detect complete and deletion HCV genomes are presented. In the lower panel, black lines indicate the full-length and deletion sequences, respectively. The shaded boxes display the position of the in-frame deletion region.

	1170	1180	1190	1200	2540	2550	2560	2570
P21 (com)	-----	-----	-----	-----	-----	-----	-----	-----
P21 (del)	CGCTATGTAC	GTGGGGGATC	TCTGCGGATC	TGTTTTCTC/GCGTCTGTGC	CTGYTTGTGG	ATGATGCTGC	TGGTAGCTCA	
	C-----C		/.....		-----C	-----R	-----R
	980	990	1000	1010	2250	2260	2270	2280
R30 (com)	-----	-----	-----	-----	-----	-----	-----	-----
R30 (del)	GCAGTATTGT	GTATGAGGCC	GATGACGTTA	TTTTGCACAC/CAAGGTGAGG	ATGTTTGTAG	GCGGTTTTGA	ACACCGTTT	
	-----G	-----T	/.....		-----G	-----G	-----A
	970	980	990	1000	2400	2410	2420	2430
D33 (com)	-----	-----	-----	-----	-----	-----	-----	-----
D33 (del)	TGCCCAAATT	CAAGTATTGT	CTATGAGGCG	GATACCATCA/TATGCCGGCT	TTGTCCACTG	GCCTGATCCA	TCTTCACCAG	
	-----A		/.....	-----A			
	900	910	920	930	2770	2780	2790	2800
D88 (com)	-----	-----	-----	-----	-----	-----	-----	-----
D88 (del)	GTGCCTTACC	ACACCTGCTT	CAGCCCTGCA	CTATGCCAAT/TTGCCCAAC	GGGCGTTCGC	GTACGATCAA	GAGCAAGCAG	
		-----Y	-----C	-----C/.....	-----C	-----T	

Figure 3. Consensus sequence alignments of different HCV full length complete genomes (com) and deletion mutants (del). Nucleotide differences between complete and deletion mutant sequences are shown in bold. Residues identical to the major sequence are indicated by a dash, the deletion region is indicated by dots and slashes indicate discontinuous sequences of the deletion regions. The nucleotide position is presented above the sequences (numbering system for the prototype strain HCV-H) and sample names are indicated in the left-hand column. Complement nucleotide codes; R = T or C, Y = G or A.

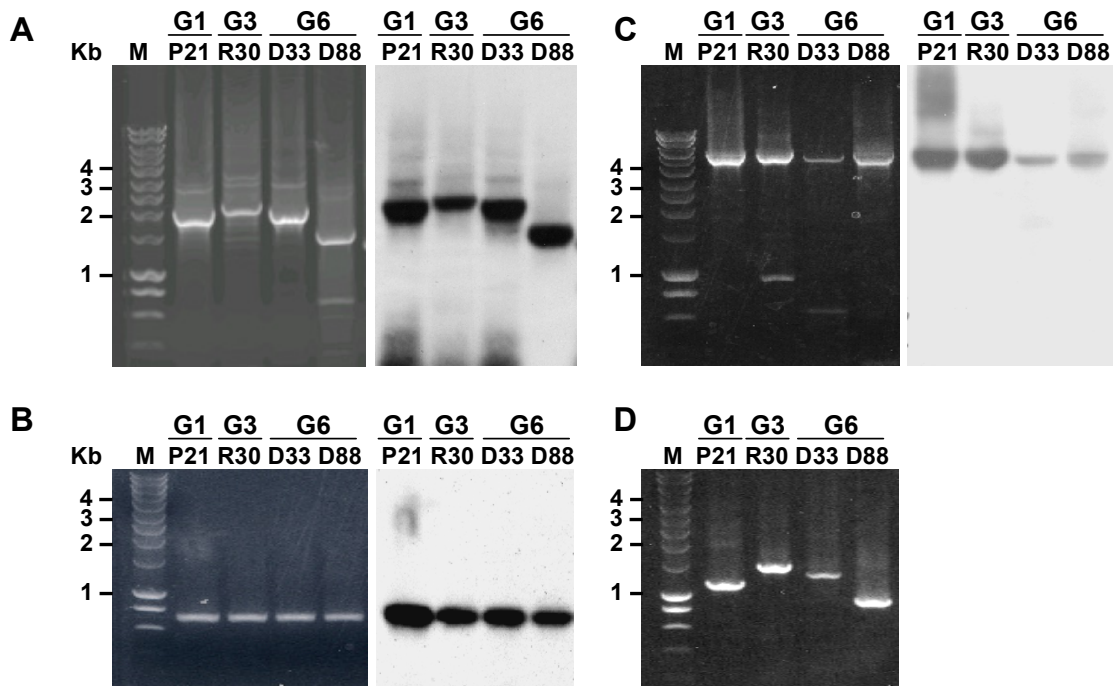


Figure 4. Experimental confirmation of HCV E1-E2 deletion mutants. RT-PCR (left panel) and southern blot analysis (right panel) of HCV genotype 1 (P21), genotype 3 (R30) and genotype 6 (D33 and D88), using sequence specific primers to amplify (a) 5'UTR to NS3 (nucleotides 17 to 3277), (b) 5'UTR to Core (nucleotides 17 to 750), and (c) NS3 to NS5 (nucleotides 4039 to 8625) regions. Amplicons were separated in 0.8% agarose gels along with a marker (M) for which the molecular weights are given on the left side in kilobasepairs (Kb). (d) Gel-electrophoresis of RT-PCR products amplified from RNA of samples P21, R30, D33, and D88 with primer 3636 and a junction site primer overlapping the deleted region that was specifically designed for each sample (Table 1). Molecular weight marker (M) is indicated on the left side.

To minimize the frequency of template switching during cDNA synthesis (18), we used (RNase H-deficient) high thermostability reverse transcriptase (RT) enzymes to perform cDNA synthesis. Using these enzymes, small sized non-specific bands disappeared, but deletion genome fragments were detected using both Transcriptor RT (Fig. 5) and Superscript III RT (not shown). Southern blot analysis confirmed the specificity of the amplified products (Fig. 5A). In addition, we designed a simulation experiment in which we performed an RT-PCR assay with *in vitro* transcribed RNA obtained from cloned DNA templates derived from the full-length or deletion genome of sample P21. Amplification of the full-length RNA template revealed only bands for full-length template (no deletion form) (Fig. 5B), whereas the template for deletion genome only showed bands for the deletion form (Fig. 5C) Based on all control experiments above, we conclude that the observed deletion mutants of HCV are bona fide natural RNA species and that the deletions are found only in the region encoding the E1, E2, p7 and NS2 proteins.

Circulating HCV deletion mutants are protected from RNase. In case that HCV deletion mutants in plasma originate from replicating viruses in the liver, viral RNAs should be packaged in viral particles and resist degradation by RNase A. Plasma was treated with Triton X-100 to disintegrate membranes and the sample was treated with RNase A. As shown in Fig. 6, similar to HCV full length genomes, HCV deletion mutants were detected in samples treated with RNase A, but not when Triton X-100 was added concomitantly. These results were confirmed using deletion specific primers (Fig. 6) Thus, circulating HCV deletion mutant genomes are most likely located within viral particles.

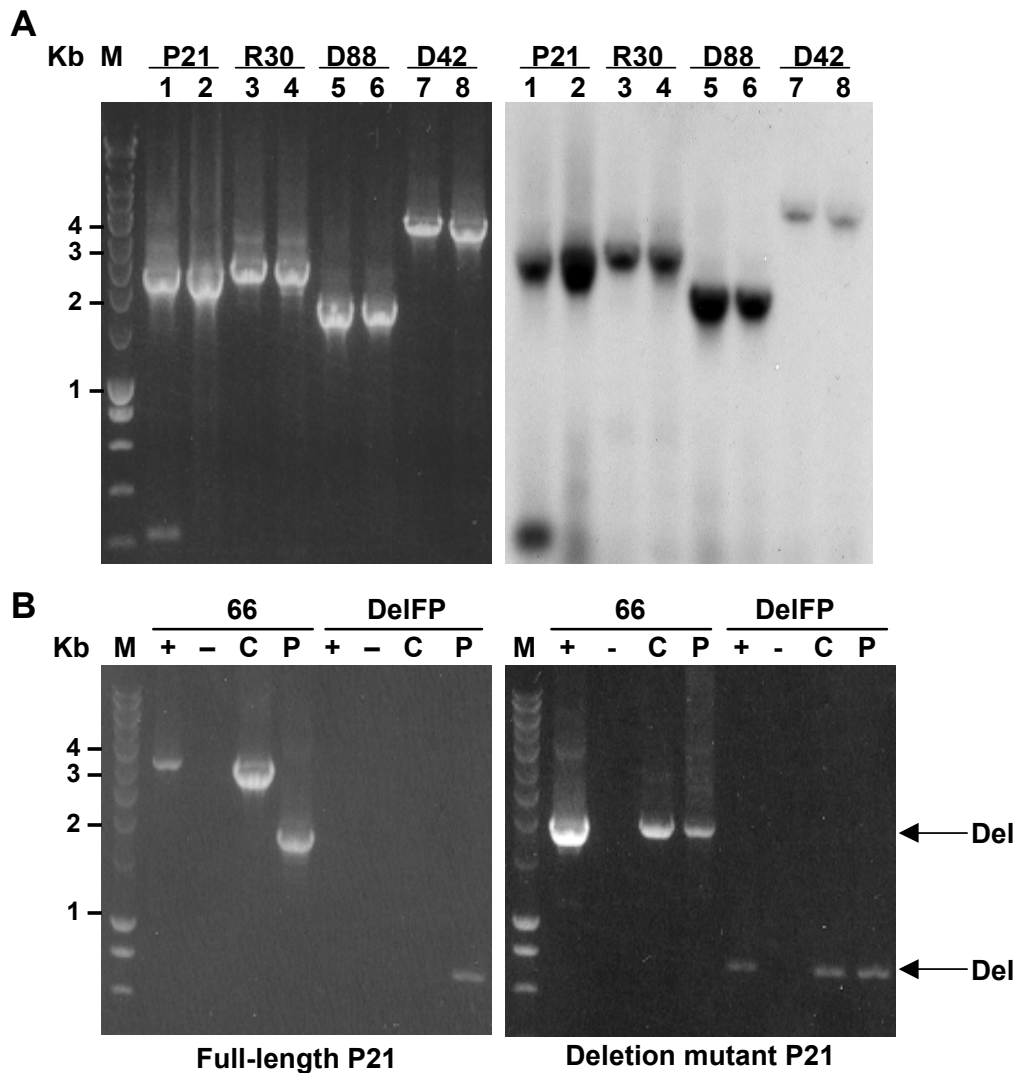


Figure 5. HCV deletion mutants are bona fide RNA species. (a) Comparison of RT-PCR amplification of HCV nucleotides 66 to 3636 (left panel) and southern blot analysis (right panel) of HCV genotype 1 (P21), genotype 3 (R30) and genotype 6 (D33 and D88) serum samples, using Expand reverse transcriptase (lanes 1, 3, 5, and 7) and Transcriptor reverse transcriptase (lanes 2, 4, 6, and 8). (b) RT-PCR analysis of *in vitro* transcribed full-length HCV RNA (left panel) and deletion mutant RNA (right panel) of sample P21 (genotype 1), using primer 3636 and either primer 66 or a junction site primer overlapping the deleted region (DelFP). Reactions were performed in the presence (+) or absence (-) of T7 RNA polymerase. As controls, a cDNA clone containing either full-length or deletion mutant HCV (C) or RNA isolated from P21 serum sample (P) were taken along. Molecular weight marker (M) is shown in kilobasepairs (Kb) on the left side. The amplicons representing deletion forms (Del) are indicated with an arrow.

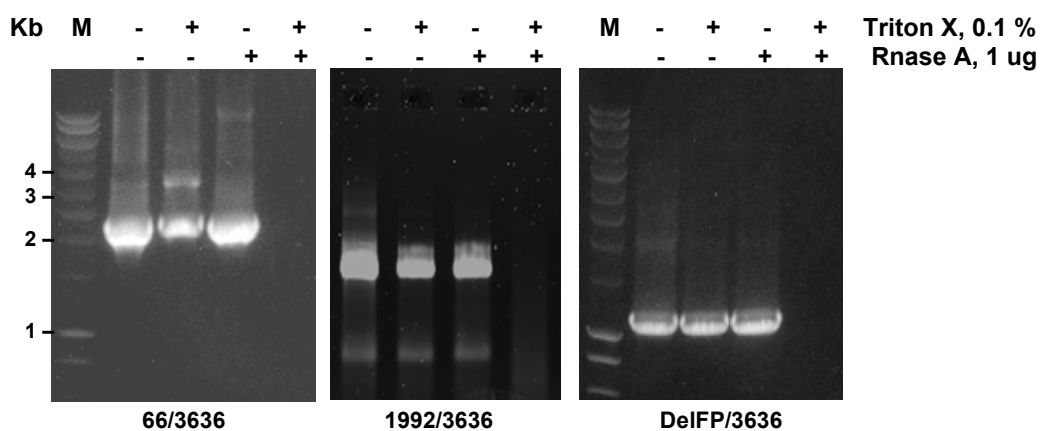


Figure 6. HCV RNA genotype 1 (P21) deletion mutant circulating in plasma is protected from RNase A digestion. The P21 serum sample was treated with or without 0.1% Triton X-100 and 1 µg RNase A before RNA extraction. RT-PCR amplifications of nucleotides 66 to 3636 (deletion mutant, left panel), 1992 to 3636 (full length, middle panel) and of the deletion mutant with a junction site primer overlapping the deleted region (DelFP) and primer 3636 (deletion mutant, right panel) are shown. Molecular weight marker (M) is indicated on the left side.

DISCUSSION

In this study, we identified and characterized circulating HCV deletion mutants in plasma of chronically infected HCV patients. These HCV RNAs always co-existed with full-length genomes, contained a large in frame deletion that varied in size from 1254 to 2022 nucleotides located in the E1-NS2 region of the genome. We concluded that these deletion mutants are natural occurring RNA species on the basis of carefully designed control experiments using deletion-specific primers for amplification, elevated temperatures for cDNA synthesis, and *in vitro* transcribed RNA as template. Moreover, samples have their own unique deletion mutant that contains nucleotide substitutions as compared to the full-length genome, suggesting independent evolution of the full-length genome and deletion mutant. Efforts to detect the full length and deletion mutants by northern blot techniques were unsuccessful. Therefore accurate estimations on the ratio of both forms present in the plasma are difficult to make. The fact that only one dominant deletion form is detected per sample, suggests that the “fittest” defective virus

is selected. Last but not least, deletion mutants likely occur in all HCV genotypes as we identified deletion mutants in ~20% of chronic HCV patients infected with genotype 1, 3 and 6. It is of note that our discovery seems to validate the claims of subgenomic HCV in liver biopsy specimens of chronic HCV patients and liver transplant recipients (12, 29). The region where the deletion occurs is similar in all studies but in contrast to our findings, these authors reported also the presence of out-frame deletion genomes. Moreover, our findings generalize the detection of HCV deletion mutants to a significant portion of the chronic hepatitis C patients infected with different genotypes.

The deletion genomes contain essential parts for autonomous HCV replication (5'UTR, core and NS2-NS5-3'UTR), regions similar to what has been described for the artificially constructed subgenomic replicons (17, 28), and may therefore represent naturally occurring replicons. Interestingly, both full-length genome and mutant genome were resistant to degradation by RNaseA, suggesting that both RNA forms are packaged in virions. Assuming that HCV particle assembly requires functional E1/E2 glycoproteins, the deletion mutants can only be packaged and secreted from cells when cells are co-infected with a wildtype virus with a full-length genome.

The HCV deletion mutants are highly reminiscent of DI particles described for many viruses (10), including members of the *Flaviviridae* family, such as tick-borne encephalitis virus, Murray Valley encephalitis virus, Dengue virus, West Nile virus and Japanese encephalitis virus (1, 3, 15, 19, 30). DI viruses are mutants that arise spontaneously when the standard wildtype virus is passaged in tissue culture at high multiplicities of infection. The DI genomes contain *cis*-acting signals required for replication, but lack part or most of the region encoding viral proteins (4). These smaller viruses can be complemented by co-infection with a helper (wild-type) virus, and might interfere with the replication of helper virus by competition for limiting factors from the host cell (16). To our knowledge, most, if not all, DI particles have been discovered in laboratory settings and HCV is the first positive-stranded RNA virus for which defective genomes have been described in a natural human infection *in vivo*.

Methods to quantify and determine the heterogeneity of HCV genomes are currently used routinely to guide antiviral therapy and to study evolution and epidemiology of this virus. The existence of deletion mutants in chronic hepatitis C patients further exemplifies the huge diversity of HCV and has some major implications.

First, viral load quantification methods based on amplification of the 5'UTR (20) detect both defective and complete genomes. Considering a high amount of circulating HCV deletion mutant, it may be of interest to differentiate between these genomic HCV forms in order to further unravel the pathogenesis of HCV and response to antiviral therapy. However, given the fact that only a fraction of the patients harbour deletion mutants a large panel of samples from well characterized patients needs to be screened. Last, but not least, it has been proposed that defective viruses may play a significant role in the establishment and maintenance of chronic infection *in vivo* (11), despite the fact that most if not all data comes from *in vitro* experiments (15, 30). Whether HCV deletion genomes play a role in the persistence of HCV remains to be delineated.

ACKNOWLEDGEMENTS

This work was supported by the Commission of the European Community (HECSA Project, ICA4-CT-1999-10009). We thank Vu Thuy Yen, Jan Brouwer, Paula van Luijt, Kavita Bedi and Apiradee Theamboonlers for technical assistance and Dr. Ron Fouchier for valuable discussion.

REFERENCES

1. **Aaskov, J., K. Buzacott, H. M. Thu, K. Lowry, and E. C. Holmes.** 2006. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* **311**:236-238.
2. **Bredenbeek, P. J., A. F. Noten, M. C. Horzinek, and W. J. Spaan.** 1990. Identification and stability of a 30-kDa nonstructural protein encoded by mRNA 2 of mouse hepatitis virus in infected cells. *Virology* **175**:303-306.
3. **Brinton, M. A.** 1983. Analysis of extracellular West Nile virus particles produced by cell cultures from genetically resistant and susceptible mice indicates enhanced amplification of defective interfering particles by resistant cultures. *J. Virol.* **46**:860-870
4. **Cann, A. J.** 2000. *RNA Viruses : a practical approach.* Series ed., B.D. Hames. Oxford University Press, Oxford.
5. **Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, R. Medina-Selby, P. J. Barr, A.J. Weiner, D.W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**:2451-2455.
6. **Fouchier, R. A., T. M. Bestebroer, S. Herfst, L. Van Der Kemp, G. F. Rimmelzwaan, and A. D. Osterhaus.** 2000. Detection of influenza A viruses from different species by PCR amplification of conserved sequences in the matrix gene. *J. Clin. Microbiol.* **38**:4096-4101.
7. **Friebe, P., and R. Bartenschlager.** 2002. Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *J. Virol.* **76**:5326-5338.
8. **Friebe, P., V. Lohmann, N. Krieger, and R. Bartenschlager.** 2001. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J. Virol.* **75**:12047-12057.
9. **Gale, M., Jr., C. M. Blakely, B. Kwieciszewski, S. L. Tan, M. Dossett, N. M. Tang, M. J. Korth, S. J. Polyak, D. R. Gretch, and M. G. Katze.** 1998. Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol. Cell. Biol.* **18**:5208-5218.
10. **Holland, J. J.** 1990. Defective viral genomes. *In* B.N. Fields, D.M. Knipe, et al. (ed.), *Virology.* Raven Press, Ltd., New York, N.Y.
11. **Huang, A. S., and D. Baltimore.** 1970. Defective viral particles and viral disease processes. *Nature* **226**:325-327.
12. **Iwai, A., H. Marusawa, Y. Takada, H. Egawa, K. Ikeda, M. Nabeshima, S. Uemoto, and T. Chiba.** 2006. Identification of novel defective HCV clones in liver transplant recipients with recurrent HCV infection. *J. Viral Hepat.* **13**:523-531.
13. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius.** 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J. Virol.* **76**:4034-4043.
14. **Kolykhalov, A. A., E. V. Agapov, K. J. Blight, K. Mihalik, S. M. Feinstone, and C. M. Rice.** 1997. Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science* **277**:570-574.

15. **Lancaster, M. U., S. I. Hodgetts, J. S. Mackenzie, and N. Urosevic.** 1998. Characterization of defective viral RNA produced during persistent infection of Vero cells with Murray Valley encephalitis virus. *J. Virol.* **72**:2474-2482.
16. **Lazzarini, R., J.D. Keene and M. Schubert.** 1981. The origins of defective interfering particles of the negative-strand RNA viruses. *Cell* **26**:145-154.
17. **Lohmann, V., F. Korner, J. Koch, U. Herian, L. Theilmann, and R. Bartenschlager.** 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**:110-113.
18. **Luo, G. and J. Taylor.** 1990. Template switching by reverse transcriptase during DNA synthesis. *J. Virol.* **64**:4321-4328.
19. **Mandl, C. W., H. Holzmann, T. Meixner, S. Rauscher, P. F. Stadler, S. L. Allison, and F. X. Heinz.** 1998. Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. *J. Virol.* **72**:2132-2140.
20. **Nolte, F. S.** 2001. Hepatitis C virus genotyping: clinical implications and methods. *Mol. Diagn.* **6**:265-277.
21. **Noppornpanth, S., T. X. Lien, Y. Poovorawan, S. L. Smits, A. D. Osterhaus, and B. L. Haagmans.** 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J. Virol.* **80**:7569-7577.
22. **Okamoto, H., M. Kojima, S. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiro.** 1992. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* **190**:894-899.
23. **Rispeter, K., M. Lu, S. Lechner, A. Zibert, and M. Roggendorf.** 1997. Cloning and characterization of a complete open reading frame of the hepatitis C virus genome in only two cDNA fragments. *J. Gen. Virol.* **78** (Pt 11):2751-2759.
24. **Shepard, C. W., L. Finelli, and M. J. Alter.** 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect. Dis.* **5**:558-567.
25. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-973.
26. **Smith, D. B., S. Pathirana, F. Davidson, E. Lawlor, J. Power, P. L. Yap, and P. Simmonds.** 1997. The origin of hepatitis C virus genotypes. *J. Gen. Virol.* **78** (Pt 2):321-328.
27. **Tseng, C. T., and G. R. Klimpel.** 2002. Binding of the hepatitis C virus envelope protein E2 to CD81 inhibits natural killer cell functions. *J. Exp. Med.* **195**:43-49.
28. **Wakita, T., T. Pietschmann, T. Kato, T. Date, M. Miyamoto, Z. Zhao, K. Murthy, A. Habermann, H. G. Krausslich, M. Mizokami, R. Bartenschlager, and T. J. Liang.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* **11**:791-796.
29. **Yagi, S., K. Mori, E. Tanaka, A. Matsumoto, F. Sunaga, K. Kiyosawa, and K. Yamaguchi.** 2005. Identification of novel HCV subgenome replicating persistently in chronic active hepatitis C patients. *J. Med. Virol.* **77**:399-413.

30. **Yoon, S. W., S. Y. Lee, S. Y. Won, S. H. Park, S. Y. Park, and Y. S. Jeong.** 2006. Characterization of homologous defective interfering RNA during persistent infection of Vero cells with Japanese encephalitis virus. *Mol. Cells* **21**:112-120.



CHAPTER 5

Complete genome analysis of Hepatitis C virus subtypes 6t and 6u

Suwanna Noppornpanth,¹ Yong Poovorawan,² Truong Xuan Lien,³ Saskia L. Smits,¹
Albert D. M. E. Osterhaus¹ and Bart L. Haagmans¹

¹*Department of Virology, Erasmus Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands*

²*Center of Excellence in Viral Hepatitis Research, Department of Pediatrics, Chulalongkorn Hospital Bangkok, Thailand*

³*Department of Biological Analysis, Pasteur Institute Ho Chi Minh City, Ho Chi Minh City, Vietnam*

Journal of General Virology, 89 (2008), p. 1276-81.

ABSTRACT

Hepatitis C virus (HCV) genomes exhibit high nucleotide sequence diversity. In this study, we performed complete genome sequence analysis of eleven HCV genotype 6 samples from Vietnam and Thailand. We identified nine HCV complete genomes belonging to subtypes 6a (D9), 6e (D42, D88), 6f (TH52), 6i (TH24), 6l (D33), 6n (TH22, TH31) and 6o (D85). Phylogenetic analysis of the Core/E1 and NS5B regions from unclassified genotype 6 isolates from Asian immigrants in Canada revealed that two other viruses (D49 and D83) could be classified as novel candidates of HCV subtypes 6t and 6u.

INTRODUCTION

Hepatitis C virus (HCV) is an enveloped positive-stranded RNA virus which belongs to the Family *Flaviviridae* and is the leading cause of acute and chronic hepatitis in about 3% of the human population (23). The entire viral genome contains about 9,600 nucleotides encoding structural (Core, E1, E2) and nonstructural (P7 and NS2-NS5B) proteins flanked by untranslated regions (5). The virus shows extreme genetic diversity and currently six major groups (genotypes 1-6) and a large number of subtypes have been identified (25). Viral genomes of different genotypes may vary by 30-35%, while nucleotide sequences of subtypes may differ by 15-20% (26).

HCV genotype 6 viruses mainly circulate in South East Asia (SEA), including Vietnam (32), Indonesia (30), Thailand (4, 29), Cambodia (3) and Myanmar (24). Currently, genotype 6 viruses are classified into 17 subtypes (6a-6q) based on complete genome analysis (9, 10, 12, 30, 31). Recently, the existence of subtype 6r and 6s viruses was proposed based on sequence analysis of core/E1 and NS5B from HCV strains isolated from Cambodian immigrants in Canada (14). In addition, a subtype of some viruses could not be assigned and these were classified as novel candidate HCV genotype 6 (8, 24).

High nucleotide sequence heterogeneity hampers complete genome characterization of new subtypes due to inefficient primer annealing. We developed techniques to obtain large RT-PCR fragments amplified from HCV positive sera using degenerate primers sets and obtained entire HCV genome sequences from 100 µl HCV positive sera (15). This led to the complete genome characterization of a Vietnamese HCV recombinant genotype 2i/6p strain (15). In this report, we selected several sera from HCV positive blood donors from Vietnam and Thailand based on a preliminary screen to include rare or unclassified HCV subtypes 6 (15) and performed full-genome sequencing to characterize further HCV subtypes 6.

MATERIALS AND METHODS

Total RNA was isolated from 100 µl serum by proteinase K/phenol chloroform extraction as described previously (18). The entire HCV genome was amplified from

three different cDNA reverse transcriptions from position 5460, 8625 and 9325 (numbering system according to prototype strain H77, Genbank accession no. AF009606). Three overlapping large PCR fragments were first-round amplified by degenerate primers, followed by nested PCR as described (15). Amplified products were gel-purified with the Qiaquick gel extraction kit (Qiagen) and ligated directly into plasmid pCR2.1 (TA cloning kit, Invitrogen). For each amplicon, five positive clones were selected and sequenced. Multiple sequence alignment was performed with the BioEdit program (version 7.0.1) according to amino acid alignment. Phylogenetic trees were constructed from Tree Puzzle 5.2 (22) and MrBayes version 3.0b4 (20) programs using genetic distance calculated with the maximum-likelihood with rate heterogeneity method and Bayesian method, respectively. Transition/transversion ratio, proportion of invariant sites and gamma distribution factors were estimated based on the real sequence datasets. Nucleotide and amino acid substitution were estimated by using HKY and WAG models respectively. Bootstrapping was calculated from 10,000 tree puzzle steps or 5000 trees generated from the MrBayes program. Pairwise nucleotide comparison was estimated based upon *p* distances with MEGA3 software (7). Phylogenetic analysis and sequence similarity test were performed using reference sequences of HCV subtype 6a-6q, published in the Los Alamos HCV database (<http://hcv.lanl.gov/content/hcv-index>). Genetic recombination events were examined by similarity plot and bootscanning analysis, using the SimPlot program (version 3.5.1 for windows98/NT/2000/XP; available from the author, Stuart C. Ray, <http://sray.med.som.jhmi.edu/SCRsoftware>).

RESULTS

We characterized 11 complete HCV genotype 6 genomes, including seven isolates from Vietnam (D9, D33, D42, D49, D83, D85, D88) and four isolates from Thailand (TH22, TH24, TH31, TH52). These genomes varied in length from 9373 to 9423 nucleotides, encoding 3016 to 3023 amino acids (Table 1).

The whole genome of TH52 shared the main branch with samples C-0044 and C-0046, representative of HCV subtype 6f (Fig. 1a). There is a high sequence similarity (94.7%) compared to the prototype strain and results from the short fragment analysis confirmed that TH52 is a member of HCV subtype 6f (Table 2 and Fig. 2). Interestingly,

Table 1 Genome organization of the new eleven HCV isolates in this study.

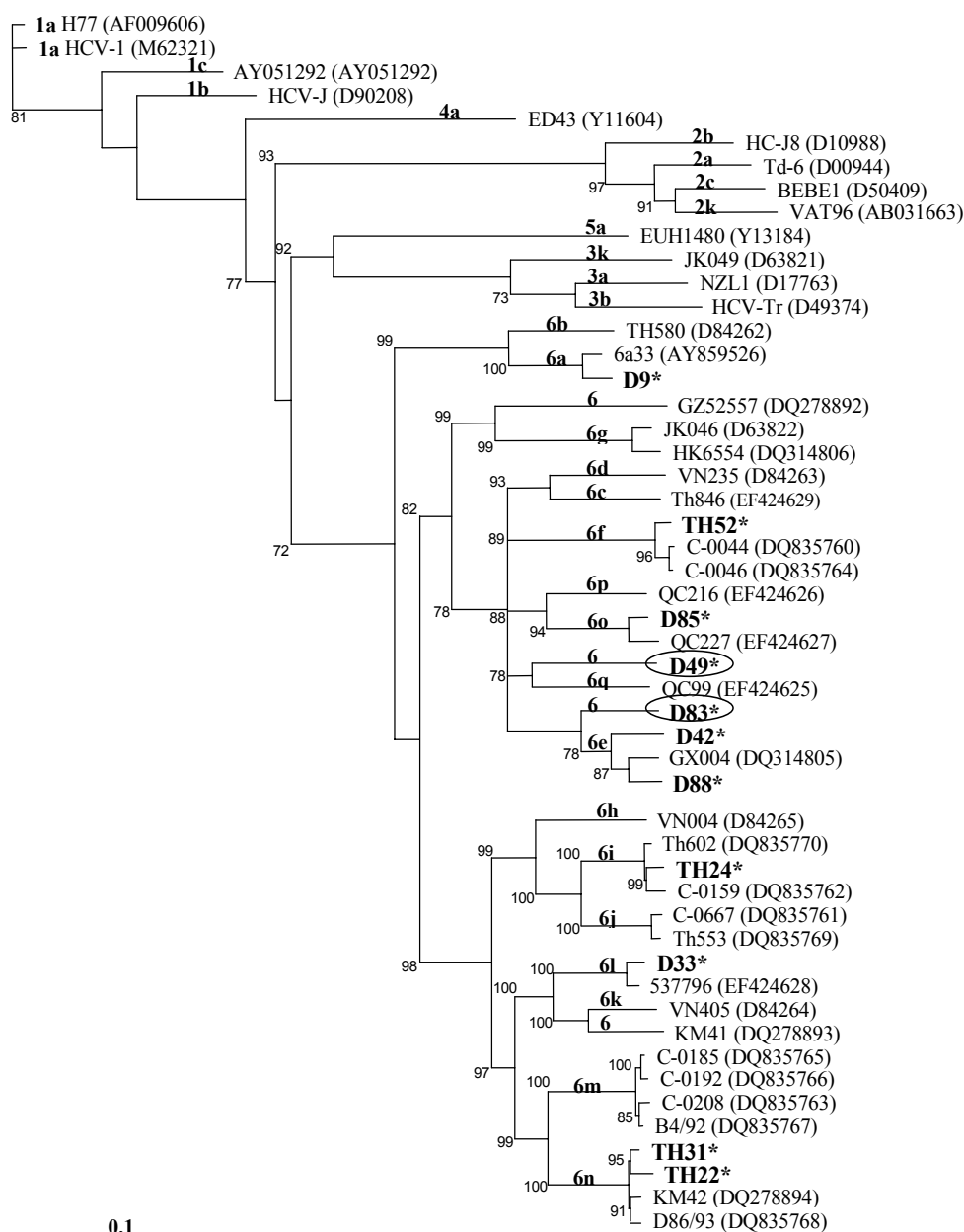
Genotype	Genome		Length of different regions (nts)												
	Strain	Nts	AAs	5'UTR	Core	E1	E2	p7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	3'UTR
6a/D9	9388	3020		328	573	576	1104	189	651	1893	162	783	1353	1776	-
6e/D42	9382	3019		325	573	576	1101	189	651	1893	162	783	1353	1776	-
6e/D88	9394	3023		325	573	576	1101	189	651	1893	162	783	1365	1776	-
6f/TH52	9382	3019		325	573	576	1089	189	651	1893	162	783	1365	1776	-
6i/TH24	9376	3017		325	573	576	1095	189	651	1893	162	783	1353	1776	-
6l/D33	9376	3017		325	573	576	1095	189	651	1893	162	783	1353	1776	-
6n/TH22	9373	3016		325	573	576	1092	189	651	1893	162	783	1353	1776	-
6n/TH31	9373	3016		325	573	576	1092	189	651	1893	162	783	1353	1776	-
6o/D85	9423	3017		325	573	576	1092	189	651	1893	162	783	1356	1776	47
6t/D49	9376	3017		325	573	576	1092	189	651	1893	162	783	1356	1776	-
6u/D83	9379	3018		325	573	576	1098	189	651	1893	162	783	1356	1776	-

Nucleotide numbering is according to reference strain H77 (GenBank accession number AF009606)
Nts= nucleotide numbers , AAs=amino acid numbers.

52/53 samples of subtype 6f classified in the Los Alamos database were obtained from Thailand (6). Full-length genomes of HCV subtype 6i (Th602 and C-0159), a subtype found in Thailand only, have been sequenced recently (10). TH24 complete genome and Core, E1 and NS5B phylogenetic trees showed 99-100% bootstrap support related to the C-0159 sequence (Fig. 1a). We identified TH22 and TH31 strains as HCV subtype 6n based on phylogenetic analysis of full-length nucleotide and amino acid sequences (Fig. 1a and 1b), while whole genome sequence similarity was 93.9 and 95.8% respectively, when compared to subtype 6n reference strain D86/93 isolated from Thailand (10). Both strains clustered with KM42 isolated from China (11) and D86/93 with high bootstrap support. HCV subtype 6n was first isolated from Thailand (2), and then later reported to be isolated from India, China and Myanmar (24).

D9 showed 92.6 % sequence similarity with subtype 6a (strain 6a33, GenBank accession no. AY859526) and 100% bootstrapping phylogeny support using complete, short fragment nucleotide and full-length amino acid sequences (Table 2, Fig. 1a, 1b and Fig. 2). HCV subtype 6a was first discovered in 1993 in Hong Kong (27) and has been distributed throughout Asia, including China (11), Vietnam (16), Taiwan and Singapore (Los Alamos HCV database).

1a)



0.1

Figure 1. Phylogenetic tree based on complete nucleotide (a) and amino acid (b) sequences. In bold, the six genotypes are indicated, subtypes are designated 1a-6q and 6 (unassigned subtype). The HCV prototype sequences are indicated by isolate name followed by the Genbank accession number in parenthesis. Confidence values (>70%) calculated from 10,000 tree puzzle steps are indicated at the major branching points. Branch lengths are drawn to scale. The eleven HCV strains determined in this study are shown in boldface and asterisks. Unassigned samples are specified with circular marks. Nucleotide sequences presented in this study were submitted to Genbank accession number EU246930- EU246940.

1b)

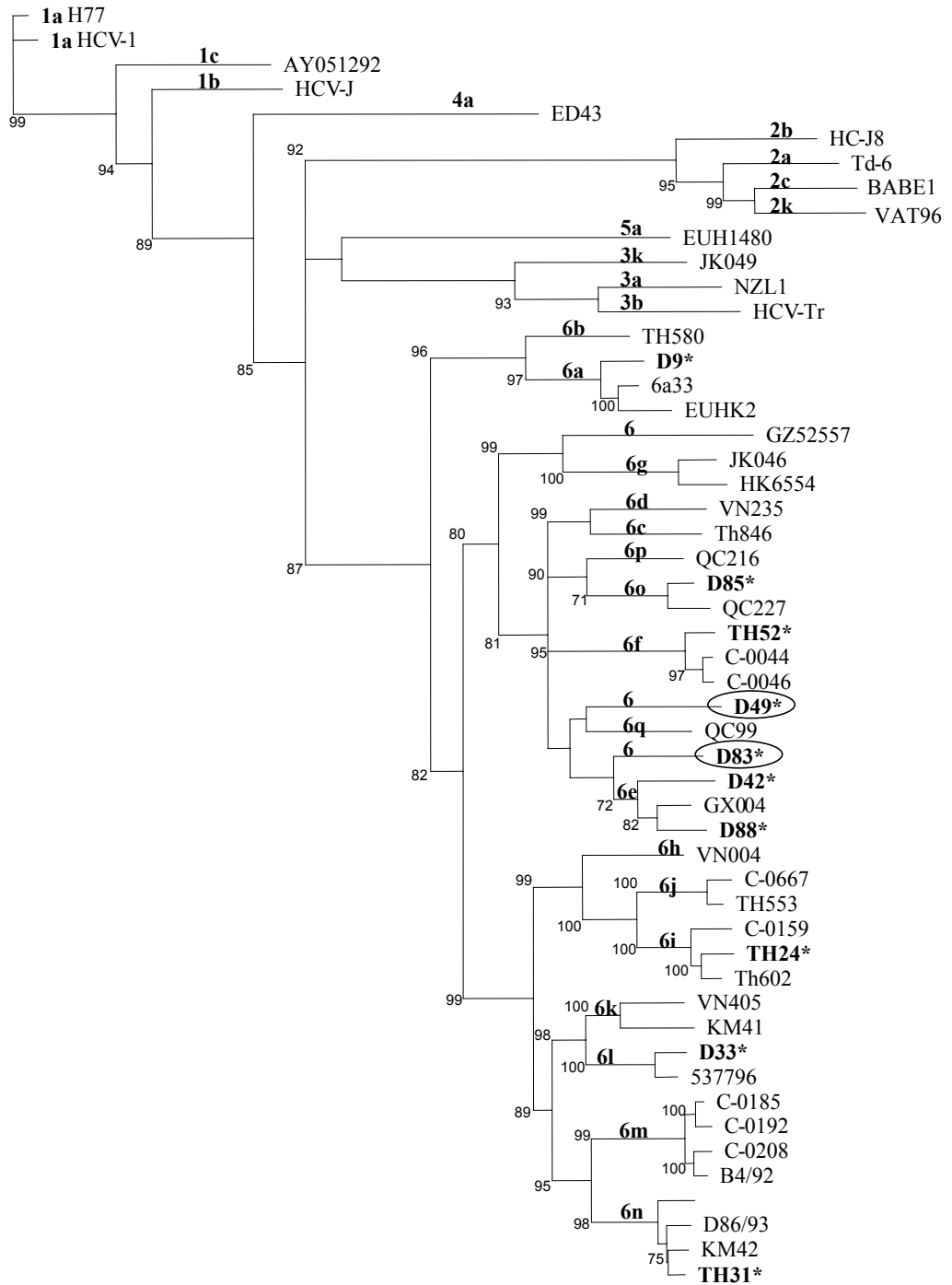


Table 2 Pairwise nucleotide similarities (%) of complete genome sequences among newly identified samples and HCV genotype 6 prototype strains.

Genotype	6a	6b	6c	6d	6e	6f	6g	6h	6i	6j	6k	6l	6m	6n	6o	6p	6q	6
Strain	6a33	TH580	TH846	VN235	GX004	C0044	JK046	VN004	Th602	Th553	VN405	537796	B4/92	D86/93	QC227	QC216	QC99	GZ52557
6a D9	92.6	79.6	74.0	73.4	73.5	73.3	73.0	72.6	73.4	73.2	73.2	73.7	73.4	73.2	74.9	74.1	74.5	73.1
6e D42	73.2	73.0	77.1	77.3	87.0	77.0	74.1	74.3	74.3	73.8	73.8	74.2	74.0	74.4	77.7	78.2	78.0	73.9
6e D88	72.9	72.5	77.7	77.4	90.8	76.9	74.6	73.3	74.0	73.7	73.6	73.9	73.8	73.7	79.1	78.1	78.5	74.0
6f TH52	73.0	72.9	77.3	77.0	77.3	94.7	74.8	73.3	73.7	73.6	73.6	74.3	74.4	73.8	77.1	78.0	77.1	74.1
6i TH24	72.8	72.3	74.0	73.5	73.5	73.1	72.8	78.9	95.5	83.0	76.5	77.2	77.5	77.2	74.1	74.7	74.7	72.7
6l D33	73.5	72.9	74.3	74.1	73.7	74.1	73.7	77.0	77.4	77.3	80.5	95.1	78.4	78.8	74.0	74.9	74.4	72.9
6n TH22	72.7	72.5	74.4	74.3	74.3	74.0	73.4	77.0	77.0	76.8	78.1	78.7	80.6	93.9	75.2	75.5	75.2	72.9
6n TH31	73.3	73.1	74.4	74.6	74.2	74.2	73.6	77.9	77.9	77.4	78.7	79.5	81.6	95.8	74.6	75.7	75.0	73.1
6o D85	73.9	73.4	78.2	77.4	78.2	77.7	74.1	73.8	74.4	73.6	73.9	74.2	74.8	74.7	92.5	79.0	80.4	74.3
6 D49	72.6	72.9	77.4	77.3	78.0	77.3	74.7	73.1	74.1	73.7	73.5	74.4	74.3	74.1	78.5	79.2	78.5	74.0
6 D83	72.7	72.8	77.4	77.6	83.6	76.9	74.5	72.9	73.8	73.6	73.0	74.1	73.5	73.8	77.1	78.6	77.9	74.2

Similarities derived from the prototype sequences are shown in bold.

Phylogenetic analysis of short regions (core, E1 and NS5B) indicated that D33 clustered to the subtype 6l prototype strain 537796 (Fig. 2). Complete genome phylogeny and pairwise similarity (95.1%) confirmed classification of D33 as subtype 6l (Fig. 1a and Table 2). Remarkably, subtype 6l and 6k were isolated from Vietnamese individuals (8, 9). D85 on the other hand had 92.5% sequence similarity to subtype 6o prototype strain QC227 isolated from an Asian immigrant in Canada (14), confirmed by phylogeny of whole genome and short fragment analysis (Fig. 1a and Fig. 2). Subtype 6o is a rare subtype in SEA and we detected only one subtype 6o virus out of 95 HCV-positive samples collected from Vietnam (data not shown).

Based on full-length nucleotide and amino acid sequences D42, D88 and D83 are closely related to subtype 6e strain GX004, recently reported from China (8) (Fig. 1a and 1b). Sequence similarities compared with GX004 were 87.0, 90.8 and 83.6 % for D42, D88 and D83, respectively (Table 2). However, phylogenetic analysis of complete genome, core and NS5B sequences revealed that D83 branched separately from other subtypes (Fig. 1a and Fig. 2). D83 was about 18% different from D42 and D88 at nucleotide and about 10% different at amino acid level (data not shown).

Based on nucleotide and amino acid sequences of the whole genome, D49 branched closely to subtype 6q (Fig. 1a and 1b). However, short genome fragment analysis using core and NS5B revealed branching with subtype 6q and 6d respectively while E1 analysis showed a separate branch linked to subtype 6p (Fig. 2). Sequence similarities compared to subtypes 6a-6q prototype sequences were 72.6-79.2% (Table 2).

Although complete genome sequences of HCV subtypes 6a-6q were used as reference strains, some genomes such as GZ52557 and KM41 isolated from China (11), and at least 23 strains (provided in the Los Alamos HCV database) based on short fragment sequences of core/E1 and NS5B (Table 3) could not be classified. Because we could not subtype D49 and D83, we re-analysed the phylogenetic tree including all subtypes (6a-6s) and unassigned subtype samples reported thus far. Now the core-E1 region of D49 showed close relation with subtype 6p and clustered with QC240, QC131 and QC145, while NS5B clustered to the same viruses but separated more clearly from other subtypes (Fig. 3a and b). Unexpectedly, both core/E1 and NS5B from D83 grouped with QC191 and QC323 viruses and clearly clustered separately from subtype 6e (Fig. 3a

and b). Thus, viruses D49, QC240, QC131 and QC145 were assigned to the novel HCV subtype 6t, while D83, QC191 and QC323 are candidate subtype 6u viruses.

2a)

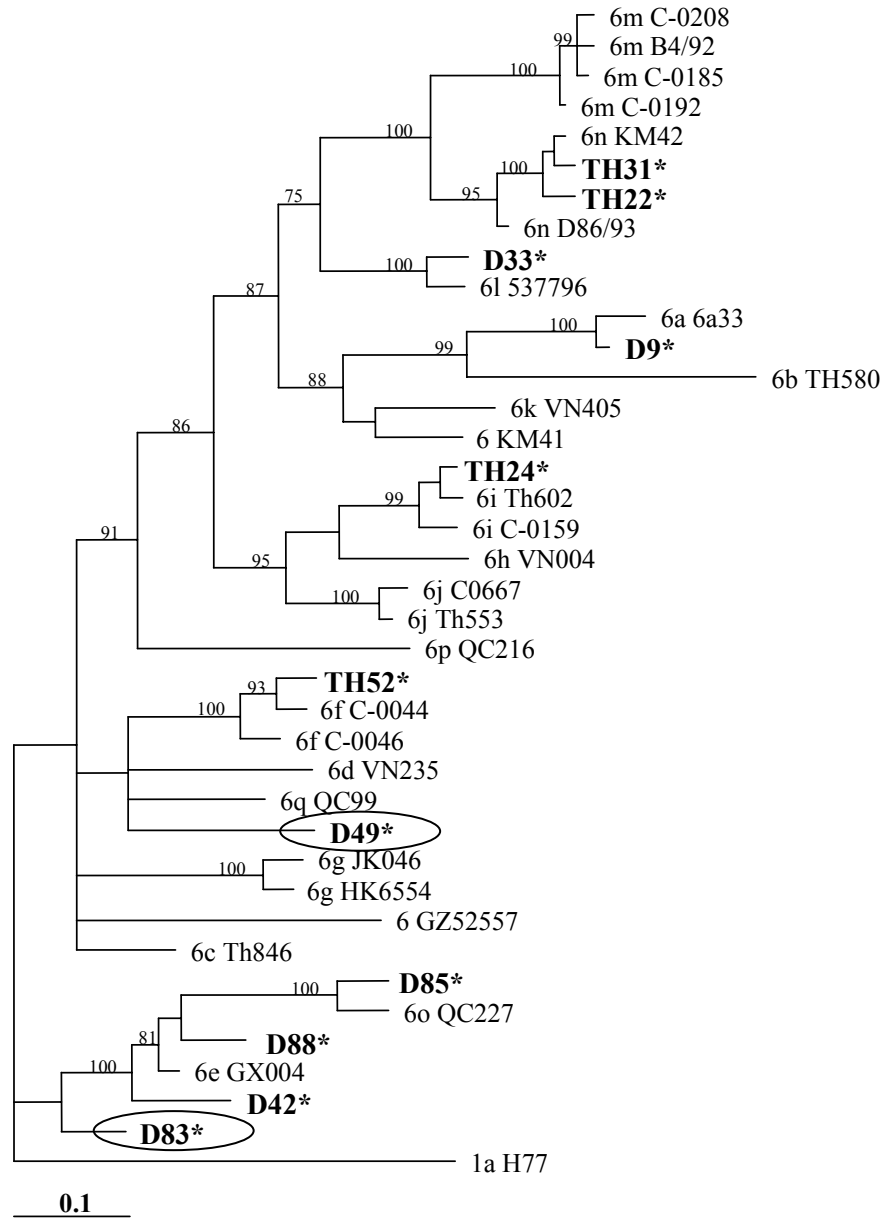
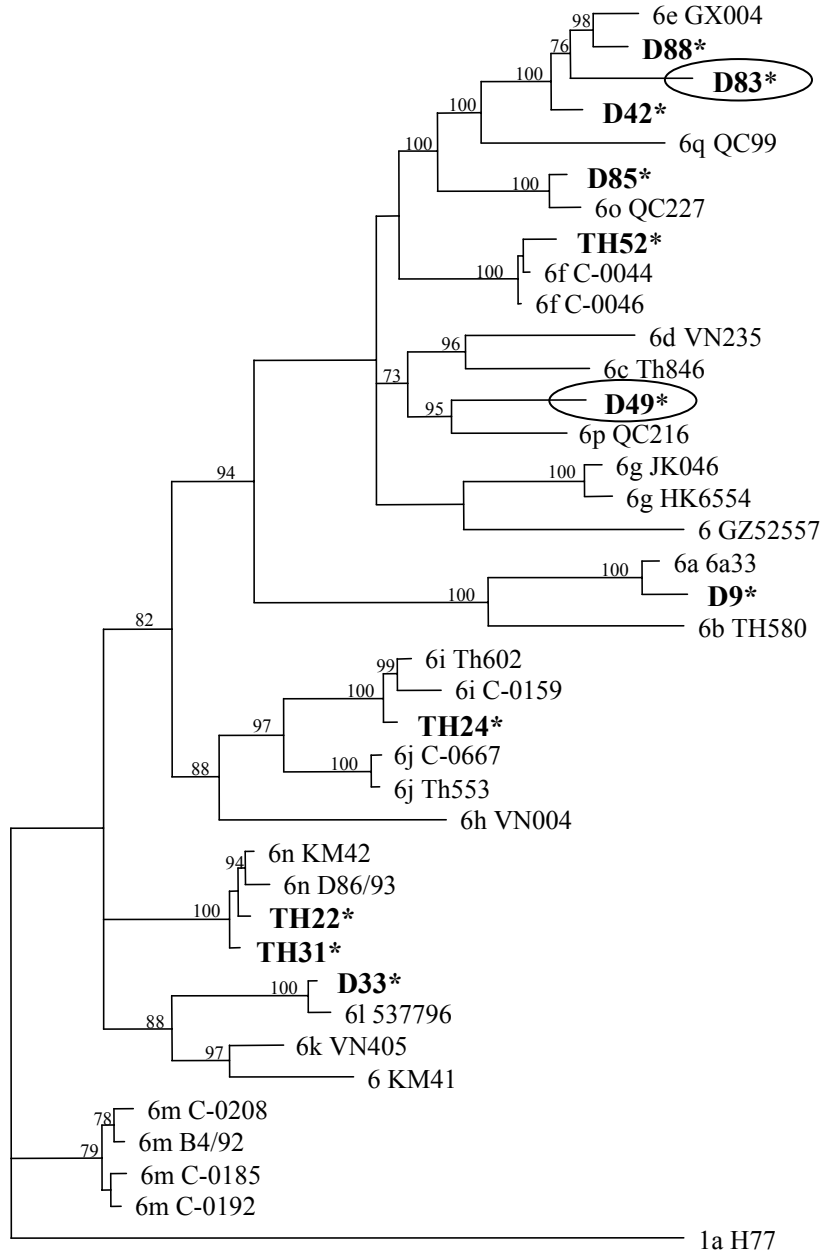


Figure. 2. Rooted Neighbor-joining trees constructed from sequences of (a) CORE (377 bps, position 342-718), (b) E1 (402 bps, position 921-1316) and (c) NS5B (324 bps, position 8282-8605) among HCV genotype 6 and outgroup sequences (H77). The HCV prototype sequences are indicated by an isolate name and subtype classification as 6a-6q and 6 (unassigned subtype). Confidence values (>70%) calculated by bootstrap analysis (1000 replicates) are indicated. Branch lengths are drawn to scale. The eleven HCV strains determined in this study are shown in bold and asterisks. Unassigned samples are specified with circular marks.

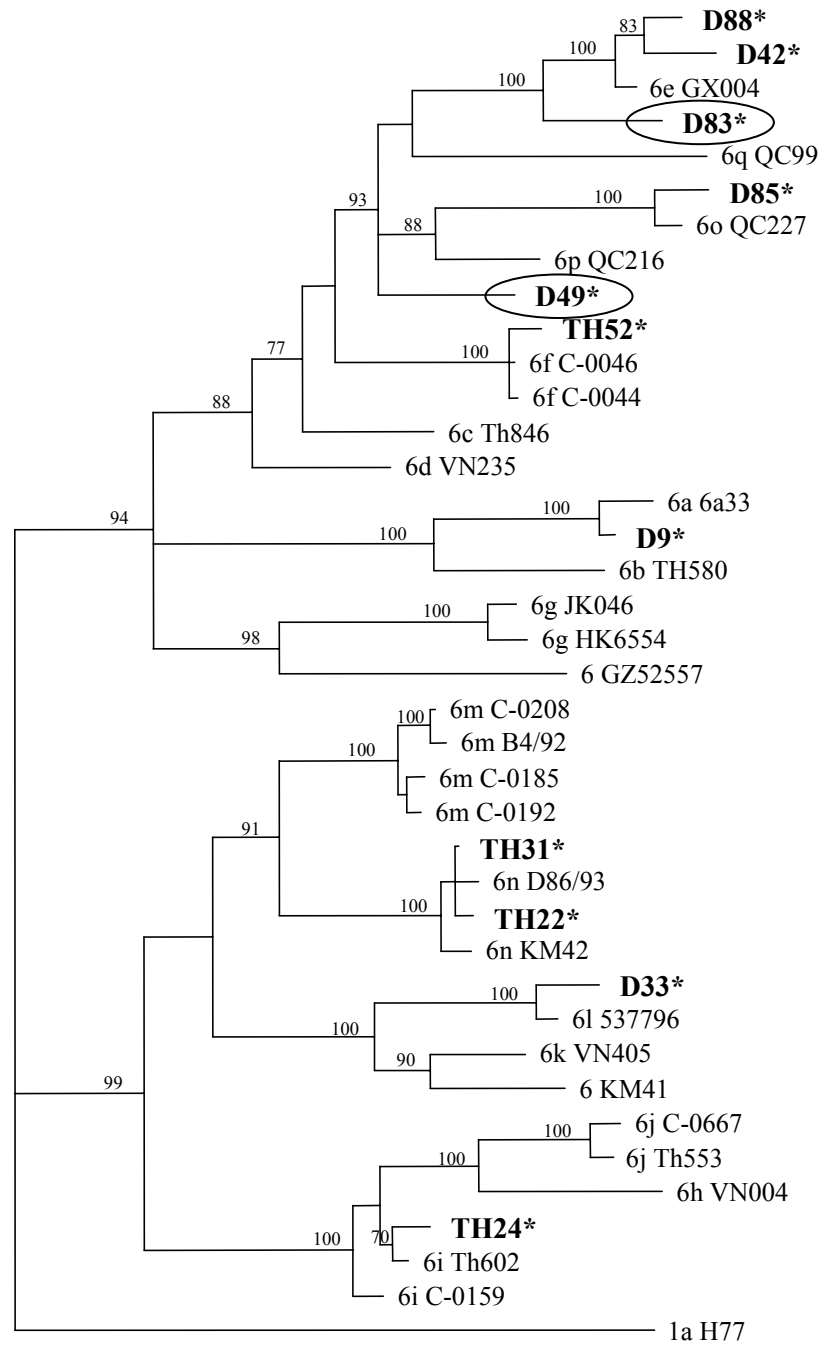
Complete genome analysis of HCV subtypes 6t and 6u

2b)



0.1

2c)



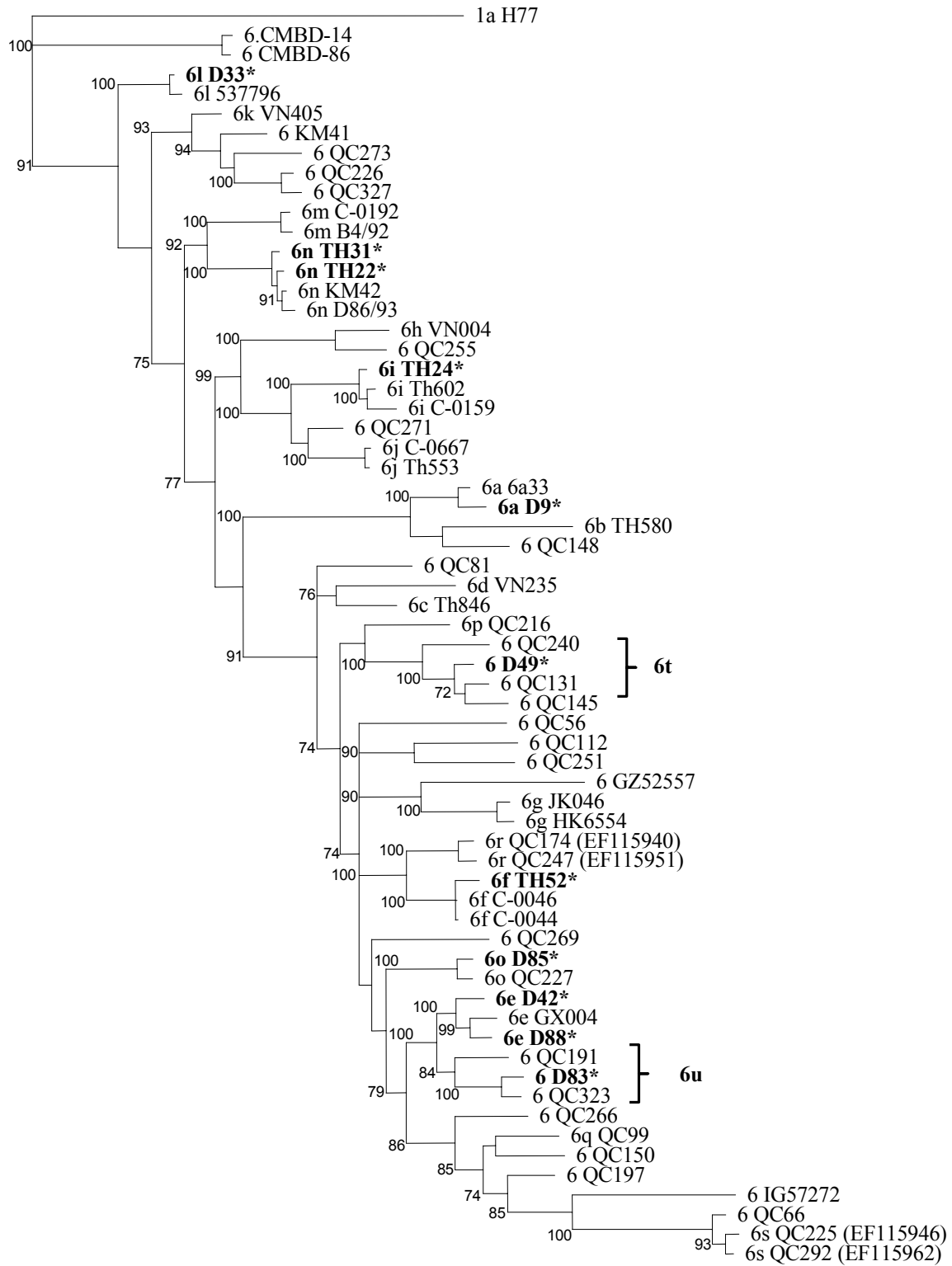
0.1

Table 3 Unassigned HCV genotype 6 strains.

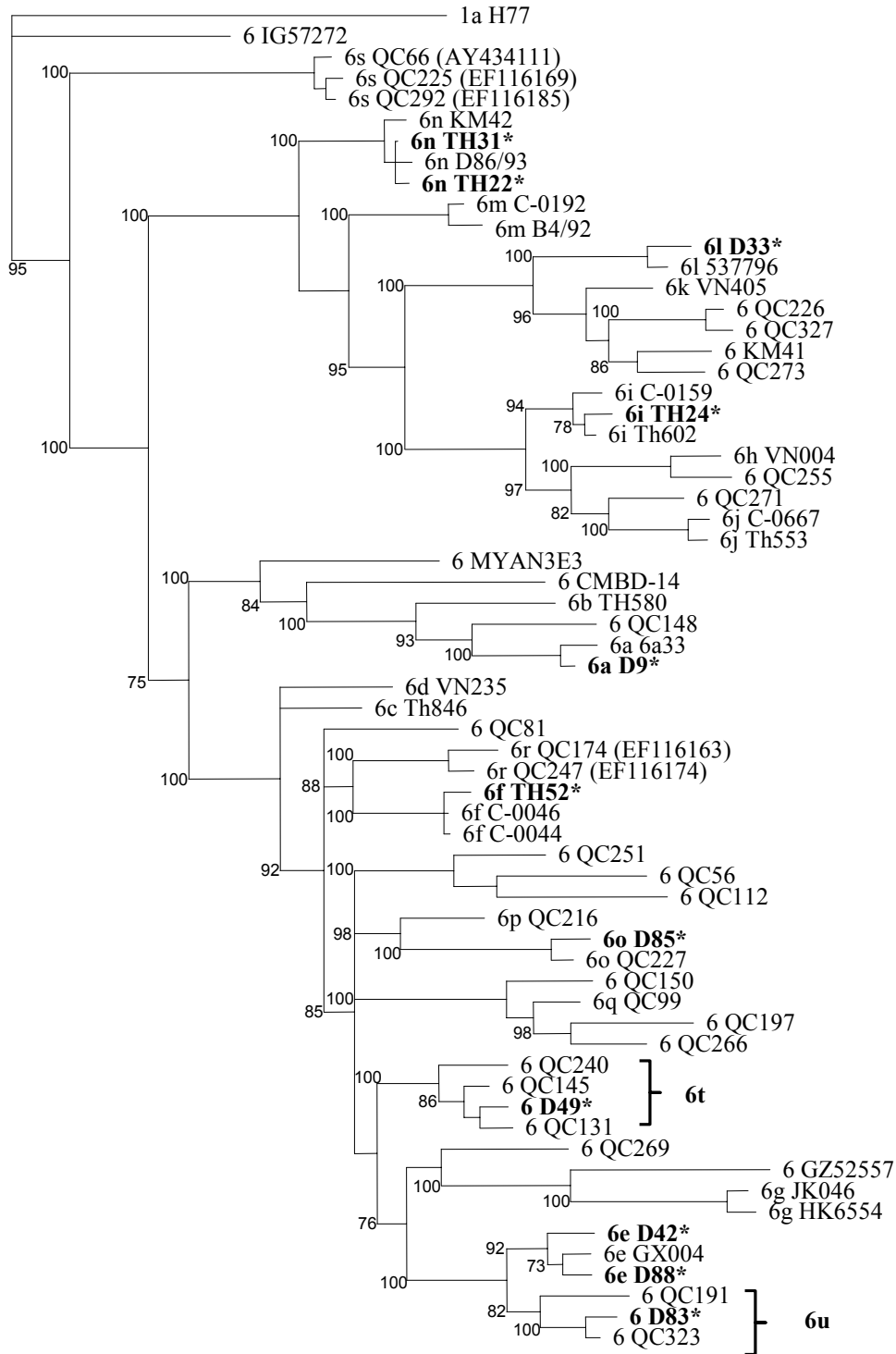
Strain name	Country	Year	Regions			
			CORE	E1	Core/E1	NS5B
KM41	China	2002	DQ278893	DQ278893	DQ278893	DQ278893
GZ52557	China	2005	DQ278892	DQ278892	DQ278892	DQ278892
IG57272	United States	2004	AY734479	AY734480	AY734479-80	AY734481
MYAN-3E-3	Myanmar	2004	-	AB254861	-	AB103143
CMBD-14	Cambodia	2005	-	DQ179116	DQ179116	DQ179117
CMBD-86	Cambodia	2005	-	DQ179143	DQ179143	-
QC56	Canada	2004	-	-	EF115927	EF116150
QC81	Canada	2004	-	-	EF115928	EF116151
QC112	Canada	2004	-	-	EF115929	EF116152
QC131	Canada	2004	-	-	EF115932	EF116155
QC145	Canada	2004	-	-	EF115934	EF116157
QC148	Canada	2004	-	-	EF115935	EF116158
QC150	Canada	2004	-	-	AY754611	AY754612
QC191	Canada	2004	-	-	EF115943	EF116166
QC197	Canada	2004	-	-	AY754629	AY754630
QC226	Canada	2004	-	-	EF115947	EF116170
QC240	Canada	2004	-	-	EF115948	EF116171
QC251	Canada	2004	-	-	EF115952	EF116175
QC255	Canada	2004	-	-	EF115953	EF116176
QC266	Canada	2004	-	-	EF115954	EF116177
QC269	Canada	2004	-	-	EF115955	EF116178
QC271	Canada	2004	-	-	EF115956	EF116179
QC273	Canada	2004	-	-	EF115957	EF116180
QC323	Canada	2004	-	-	EF115974	EF116197
QC327	Canada	2004	-	-	EF115975	EF116198

Figure 3. Rooted neighbor-joining trees depicting the phylogenetic relationship among HCV genotype 6 unassigned sequences and prototype strains. Trees were constructed for (a) a 424-nucleotide segment of the CORE/E1 genes (position 869-1292) and (b) a 324-nucleotide segment of the NS5B gene (position 8282-8605). Bootstrap values (5000 replicates) are shown along each main branch. Branch lengths are drawn to scale. The prototype HCV genotype 6 strains, indicated by an isolated name, are designated 6a-6s and 6 (unassigned subtype). Genbank Accession numbers of subtype 6s and 6r were indicated in parenthesis after the isolate name. HCV genotype 1a (H77, accession number AF009606) was used as an outgroup sequence. The eleven HCV strains obtained in this study are shown in bold and asterisks. The new assigned subtypes 6t and 6u are specified with parenthesis marks.

3A)



3B)



0.1

DISCUSSION

Using full-length genome sequences is considered the gold standard of HCV genotype/subtype identification. Due to the heterogeneity, whole genome amplification using specific primers is troublesome and 25-30 amplicons are required (12). We used degenerate primers to amplify a large PCR fragment and generated three to five amplicons, resulting in cost and time efficient characterization of the HCV genomes. Based on the pairwise sequence similarity and phylogenetic analysis of full-length and short fragment sequences, we identified complete genomes of HCV subtype 6a (D9), 6e (D42, D88), 6f (TH52), 6i (TH24), 6l (D33), 6n (TH22 and TH31) and 6o (D85). Moreover, two new subtypes 6t (D49) and 6u (D83) were assigned. Recombination between the two newly described genomes and other HCV sequences was not observed (data not shown).

Although phylogenetic analysis based on the full-length genome is the most accurate method for HCV classification, most HCVs have been classified based on short sequences analysis of core, E1 and NS5B regions (19, 21, 26). Historically, HCV classification based on short sequence phylogenetic analysis using a simple nucleotide substitution model has led to incorrect branching between closely related subtypes. HCV genotype 7, 8 and 9 from SEA have been identified using the unweighted pair-group method with arithmetic mean (UPGMA) for phylogenetic tree construction (32). However, combining the complex nucleotide substitution model with neighbor-joining results in grouping of these sequences to HCV genotype 6 subtypes (13, 28). In our study, we performed the classification based on the HKY nucleotide substitution model and maximum-likelihood method to increase the accuracy of nucleotide phylogenetic analysis. Moreover, we generated protein phylogenetic trees based on amino acid sequences to confirm our results.

Even though a range of HCV subtype 6a-6q whole genomes has been characterized, a number of viruses cannot be classified (14). These unassigned samples may represent candidate new HCV subtypes. Assignment of subtype 6t and 6u in our study is in line with the HCV genotyping consensus proposal (26). NS5B and core/E1 sequences of at least three examples are required for a provisional designation of HCV new subtype and at least 15% difference at the nucleotide level to other HCV subtypes

should be present (26). We have defined two full-length genome sequences that represent new subtypes of HCV genotype 6. Subtype 6t comprises D49, QC240, QC131 and QC145, all isolated from Vietnamese individuals and subtype 6u includes D83, which grouped to QC191 and QC323 and shows a separate branch from subtype 6e. Sample QC191 and QC323 originated from Asian immigrants in Canada but remained unassigned with respect to subtype classification (14). Grouping of D83 sequence obtained from our study with these strains suggests that subtype 6u circulates in SEA. High sequence similarity of HCV subtypes 6e and 6u indicates that both may share the same origin but have shown genetic drift as a result of long-term circulation.

Molecular evolution of HCV genotype 6 based on phylogenetic relationship separates all subtypes into three groups. Subtypes 6a and 6b are related to each other, subtypes 6c, 6d, 6e, 6f, 6g, 6p, 6q and 6o share the same phylogenetic branch and the last group contains subtypes 6h, 6i, 6j, 6k, 6l, 6m and 6n. Epidemiologic studies of HCV genotype 6 have revealed that certain subtypes, such as 6d, 6h and 6l are found in Vietnam while 6c, 6f, 6i and 6j are found in Thailand, whereas subtype 6m and 6n were isolated mainly from Thailand and Myanmar. Although there may be some sampling bias and information about subtype distribution is lacking from some countries such as Laos, Cambodia and Myanmar, these results suggest that spread of HCV occurs in closed populations. These data are consistent with the idea that HCV genome diversity is related to geographic distribution and routes of viral transmission in certain areas. Whereas genotype 1, 2 and 3 have spread worldwide mainly through contaminated blood and intravenous drug users (1), the genotype 4, 5 and 6 viruses are more restricted to African and Asian countries (26) and vary genetically as a result of long endemicity in certain areas (17).

ACKNOWLEDGEMENTS

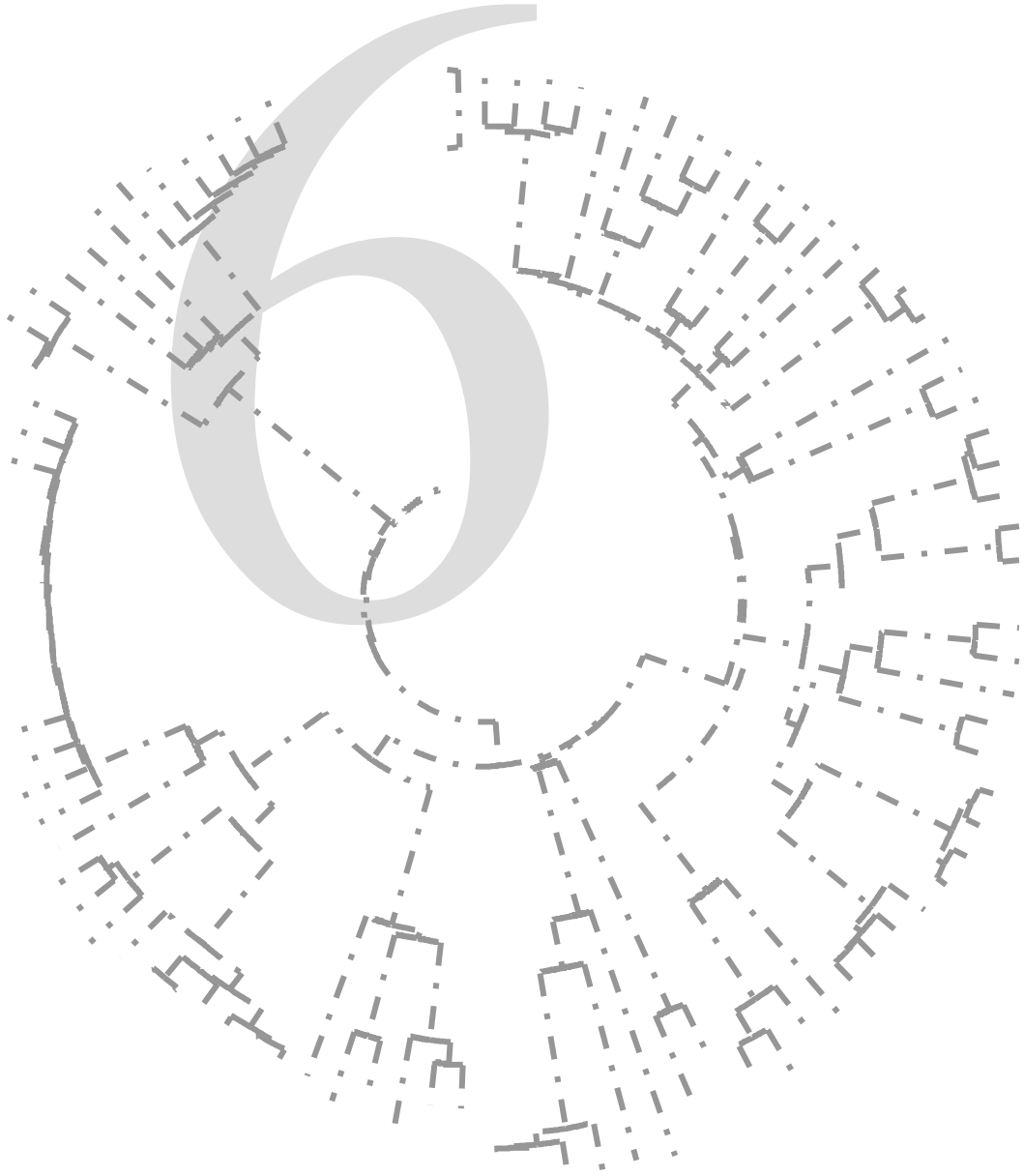
This work was supported by Commission of the European Community (HECSA Project ICA4-CT-1999-10009). We would like to thank you Dr. Vu Thuy Yen, Jan Brouwer, Paula van Luijt, Apiradee Theamboonlers and Kavita Bedi for their technical assistance.

REFERENCES

1. **Alter, M. J.** 1999. Hepatitis C virus infection in the United States. *J Hepatol* **31 Suppl 1**:88-91.
2. **Apichartpiyakul, C., C. Chittivudikarn, H. Miyajima, M. Homma, and H. Hotta.** 1994. Analysis of hepatitis C virus isolates among healthy blood donors and drug addicts in Chiang Mai, Thailand. *J Clin Microbiol* **32**:2276-9.
3. **Caruana, S. R., H. A. Kelly, S. L. De Silva, L. Chea, S. Nuon, P. Saykao, N. Bak, and B. A. Biggs.** 2005. Knowledge about hepatitis and previous exposure to hepatitis viruses in immigrants and refugees from the Mekong Region. *Aust N Z J Public Health* **29**:64-8.
4. **Chinchai, T., J. Labout, S. Noppornpanth, A. Theamboonlers, B. L. Haagmans, A. D. Osterhaus, and Y. Poovorawan.** 2003. Comparative study of different methods to genotype hepatitis C virus type 6 variants. *J Virol Methods* **109**:195-201.
5. **Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, R. Medina-Selby, P. J. Barr, and et al.** 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A* **88**:2451-5.
6. **Kuiken, C., K. Yusim, L. Boykin, and R. Richardson.** 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics* **21**:379-84.
7. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**:150-63.
8. **Li, C., Y. Fu, L. Lu, W. Ji, J. Yu, C. H. Hagedorn, and L. Zhang.** 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**:1061-9.
9. **Lu, L., C. Li, Y. Fu, F. Gao, O. G. Pybus, K. Abe, H. Okamoto, C. H. Hagedorn, and D. Murphy.** 2007. Complete genomes of hepatitis C virus (HCV) subtypes 6c, 6l, 6o, 6p and 6q: completion of a full panel of genomes for HCV genotype 6. *J Gen Virol* **88**:1519-25.
10. **Lu, L., C. Li, Y. Fu, L. Thaikruea, S. Thongsawat, N. Maneekarn, C. Apichartpiyakul, H. Hotta, H. Okamoto, D. Netski, O. G. Pybus, D. Murphy, C. H. Hagedorn, and K. E. Nelson.** 2007. Complete genomes for hepatitis C virus subtypes 6f, 6i, 6j and 6m: viral genetic diversity among Thai blood donors and infected spouses. *J Gen Virol* **88**:1505-18.
11. **Lu, L., T. Nakano, Y. He, Y. Fu, C. H. Hagedorn, and B. H. Robertson.** 2005. Hepatitis C virus genotype distribution in China: predominance of closely related subtype 1b isolates and existence of new genotype 6 variants. *J Med Virol* **75**:538-49.
12. **Lu, L., T. Nakano, C. Li, Y. Fu, S. Miller, C. Kuiken, B. H. Robertson, and C. H. Hagedorn.** 2006. Hepatitis C virus complete genome sequences identified from China representing subtypes 6k and 6n and a novel, as yet unassigned subtype within genotype 6. *J Gen Virol* **87**:629-34.
13. **Mizokami, M., T. Gojobori, K. Ohba, K. Ikeo, X. M. Ge, T. Ohno, E. Orito, and J. Y. Lau.** 1996. Hepatitis C virus types 7, 8 and 9 should be classified as type 6 subtypes. *J Hepatol* **24**:622-4.

14. **Murphy, D. G., B. Willems, M. Deschenes, N. Hilzenrat, R. Mousseau, and S. Sabbah.** 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**:1102-12.
15. **Noppornpanth, S., T. X. Lien, Y. Poovorawan, S. L. Smits, A. D. Osterhaus, and B. L. Haagmans.** 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J Virol* **80**:7569-77.
16. **Noppornpanth, S., E. Sablon, K. De Nys, X. L. Truong, J. Brouwer, M. Van Brussel, S. L. Smits, Y. Poovorawan, A. D. Osterhaus, and B. L. Haagmans.** 2006. Genotyping hepatitis C viruses from Southeast Asia by a novel line probe assay that simultaneously detects core and 5' untranslated regions. *J Clin Microbiol* **44**:3969-74.
17. **Pybus, O. G., M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey.** 2001. The epidemic behavior of the hepatitis C virus. *Science* **292**:2323-5.
18. **Rispeter, K., M. Lu, S. Lechner, A. Zibert, and M. Roggendorf.** 1997. Cloning and characterization of a complete open reading frame of the hepatitis C virus genome in only two cDNA fragments. *J Gen Virol* **78 (Pt 11)**:2751-9.
19. **Robertson, B., G. Myers, C. Howard, T. Bretin, J. Bukh, B. Gaschen, T. Gojobori, G. Maertens, M. Mizokami, O. Nainan, S. Netesov, K. Nishioka, T. Shin i, P. Simmonds, D. Smith, L. Stuyver, and A. Weiner.** 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**:2493-503.
20. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572-4.
21. **Sandres-Saune, K., P. Deny, C. Pasquier, V. Thibaut, G. Duverlie, and J. Izopet.** 2003. Determining hepatitis C genotype by analyzing the sequence of the NS5b region. *J Virol Methods* **109**:187-93.
22. **Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler.** 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**:502-4.
23. **Shepard, C. W., L. Finelli, and M. J. Alter.** 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* **5**:558-67.
24. **Shinji, T., Y. Y. Kyaw, K. Gokan, Y. Tanaka, K. Ochi, N. Kusano, T. Mizushima, S. Fujioka, H. Shiraha, A. A. Lwin, Y. Shiratori, M. Mizokami, M. Khin, M. Miyahara, S. Okada, and N. Koide.** 2004. Analysis of HCV genotypes from blood donors shows three new HCV type 6 subgroups exist in Myanmar. *Acta Med Okayama* **58**:135-42.
25. **Simmonds, P.** 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol* **85**:3173-88.
26. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-73.

27. **Simmonds, P., E. C. Holmes, T. A. Cha, S. W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg, and M. S. Urdea.** 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.
28. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 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.
29. **Thaikrua, L., S. Thongsawat, N. Maneekarn, D. Netski, D. L. Thomas, and K. E. Nelson.** 2004. Risk factors for hepatitis C virus infection among blood donors in northern Thailand. *Transfusion* **44:**1433-40.
30. **Tokita, H., H. Okamoto, H. Iizuka, J. Kishimoto, F. Tsuda, L. A. Lesmana, Y. Miyakawa, and M. Mayumi.** 1996. Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups. *J Gen Virol* **77 (Pt 2):**293-301.
31. **Tokita, H., H. Okamoto, H. Iizuka, J. Kishimoto, F. Tsuda, Y. Miyakawa, and M. Mayumi.** 1998. The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9 and comparison with those in the other eight genetic groups. *J Gen Virol* **79 (Pt 8):**1847-57.
32. **Tokita, H., H. Okamoto, F. Tsuda, P. Song, S. Nakata, T. Chosa, H. Iizuka, S. Mishiro, Y. Miyakawa, and M. Mayumi.** 1994. Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *Proc Natl Acad Sci U S A* **91:**11022-6.



CHAPTER 6

Molecular evolution of Hepatitis C Viruses: dating the origin of viruses using phylogenetic and relaxed molecular clock analysis.

Suwanna Noppornpanth¹, Byron Martina¹, Yong Poovorawan², Truong Xuan Lien³, Saskia L Smits¹, Albert DME Osterhaus¹, and Bart L Haagmans^{1*}

¹*Department of Virology, Erasmus Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands*

²*Center of Excellence in Viral Hepatitis Research, Department of Pediatrics, Chulalongkorn Hospital Bangkok, Thailand*

³*Department of Biological Analysis, Pasteur Institute Ho Chi Minh City, Ho Chi Minh City, Vietnam*

Manuscript in preparation

ABSTRACT

Estimating the evolutionary rate and time of Hepatitis C virus (HCV) divergence in humans has been proven daunting. In this study, we investigated the evolutionary pattern of HCV including analysis of phylogenetic informative regions, best-fitting nucleotide substitution model, phylogeny reconstruction and molecular clock hypothesis testing by analysing the entire genome sequences of HCV genotype 1-7 as well as individual genomic regions. Our results show that phylogenetic trees of complete genome and NS3 sequences, created under the GTR+I+ Γ model can be used for accurate HCV classification. Evolutionary rate and molecular clock hypothesis testing have been addressed by a Bayesian Markov Chain Monte Carlo (MCMC) method, accounting for lineage-specific rate variation and using a relaxed molecular clock. Our analysis revealed a relative high average rate of evolution based on the full-length genome (5.71×10^{-4} substitution/site/year), which was 2.5 and 3.3 times higher in the NS3 and NS5A gene respectively. The approximated age of the entire tree was about 4600 years, based on the whole genome sequences. The estimated time to the most recent common ancestor (tMRCA) of the different genotypes suggest the emergence of genotype 2 and 7 about 4400 years ago, of genotype 3 and 5 around 4000 years ago, and of genotype 1 and 4 about 3000 years ago. Genotype 6 represents the youngest clade, which evolved about 2600 years ago.

INTRODUCTION

Hepatitis C Virus (HCV) is a positive strand RNA virus classified as member of the genus *Hepacivirus*, family *Flaviviridae* (27). Its complete genome organization, resembles that of the pestiviruses and flaviviruses and encodes untranslated regions (5'UTR and 3'UTR), structural (Core, E1, E2) and nonstructural (p7, NS2-NS5) proteins (5, 8, 9). HCV is an important pathogen that causes chronic hepatitis, cirrhosis and hepatocellular carcinoma. WHO estimates that 3% of the world's population has been infected with HCV and there are more than 170 million chronic carriers who are at risk of developing liver cirrhosis and/or liver cancer (40).

HCV evolution is believed to be mediated by several processes such as genetic drift caused by lack of proof-reading activity in the viral RNA-dependent RNA-polymerase during replication (36), immune pressure (2) and recombination (12). Based on the genetic distance among isolates, HCV is classified into six major genotypes (genotype 1-6) with a huge number of subtypes. Genotypes may vary by 30-35% of their viral genomes, while subtypes may differ by 15-20% of their nucleotide sequences (31). Genotyping of HCV is clinically important in determining potential response to interferon therapy and the required duration of such therapy. For instance, genotypes 1 and 4 are less responsive to interferon-based treatment and require prolonged therapeutic intervention compared to genotypes 2, 3, 5 and 6 (15). In general, HCV genotypes are associated with a particular geographical region (1, 27, 3, 4, 16, 6, 17, 28). Dating the origin of the HCV is difficult because of the lack of HCV or HCV-like variants occurring in ape or monkey species (14). In theory, it may be possible to calculate the times of divergence of the six main clades and also the time of splitting of HCV subtypes by using constant nucleotide substitution rate over time. Based on this strategy, Smith et al. calculated the origin of HCV based on analysis of the E1 and NS5B genes, suggesting that the subtypes diverged around 300 years ago, and the divergence of the different genotypes should have occurred between 500 and 2000 years ago (33).

The aim of this study was to use statistical testing in order to establish the best-fit model of evolution using different data sets. We evaluated phylogenies derived from whole genome HCV sequences and several genomic regions of HCV, and employed a

Bayesian Markov Chain Monte Carlo (MCMC) approach to address phylogenetic relationships and dating the origin of genotypes.

MATERIALS AND METHODS

Compilation of sequence data.

Data sets of full-length genomes, representing all HCV genotypes, were downloaded from Los Alamos Hepatitis C Virus Database (13), <http://hcv.lanl.gov/content/hcv-index>). Only sequences were used for which the exact year and place of isolation were available. Recombinant sequences were excluded. A data set of 96 complete genomes belonging to genotype 1-7 was compiled (Table 1). The coding nucleotide sequences were aligned in-frame, using the ClustalW algorithm implemented in Bioedit Program (version 7.0.9). All sequence alignments are available from the authors on request.

Analysis of the phylogenetic signal/noise.

The presence of phylogenetic signal across the genome was examined by performing a likelihood mapping analysis on groups of four sequences (quartets) randomly chosen (35). Likelihood mapping analyses were performed with the program TREE-PUZZLE (26) on the HCV full-length genome and different genomic regions (including UTRs). For each analysis all 10,000 possible quartets for the 96 sequences of all genotypes were evaluated.

Models of nucleotide substitution and likelihood ratio test.

The best-fit substitution model was determined by MODELTEST 3.06 (http://darwin.uvigo.es/software/modeltest_server.html) (22), and likelihood scores were compared using the hierarchical likelihood ratio tests (hLRT) and the Akaike Information Criterion (AIC) implemented in PAUP version 4.0b10 (38).

Best phylogenetic tree search using likelihood Ratchets and PAUP strategies.

PAUP 4.0b10 was used to infer maximum likelihood (ML) phylogenetic trees generated with the best-fitting nucleotide substitution model. Accelerated likelihood explorations were performed using likelihood ratchets (39) in conjunction with PAUPRat (29) for more efficient searching of the likelihood surface. Likelihood ratchet uses multiple sequential truncated searches on different starting trees created by fast

algorithmic searches on reweighed data, analogous to the parsimony ratchet (20), which has been shown to increase search efficiency (39). Best trees of HCV full-length genome and regions that contain high phylogenetic signal were constructed from at least 10 iterations of the likelihood ratchets by using Heuristic ML tree search parameter with Branch-swapping subtree-pruning-regrafting (SPR) algorithm. Best trees generated from each iterations was compared and consensus tree was performed by PAUP program.

Evolutionary analysis.

Overall rates of evolutionary change (nucleotide substitutions per site per year) and the age to the most recent common ancestor (tMRCA) for each branch (genotypes) of the tree were estimated using the Bayesian MCMC approach available in the Bayesian evolutionary analysis by sampling trees (BEAST) package (http://beast.bio.ed.ac.uk/Main_Page, version 1.4.7) (7). HCV sequences recovered from the same epidemic area or isolated from nonhuman primates were excluded from the analyses. After removing all gaps, the 42 HCV genomes were used for analysis (Table 1). Various parameters such as gamma distribution (Γ 4), invariant sites (I), specific partition into codon position, i.e., 2 partitions (codon positions (1+2)) and 3 partitions (codon positions 1,2,3) were applied to test the best evolutionary model. Data sets were analysed using a constant population size under both strict and relaxed molecular clock (uncorrelated lognormal) models, which do not assume a constant rate across lineages for molecular evolution study (24). In each case, MCMC chains were run for at least 10 million times to achieve the Effective Sample Sites (ESS), and models were compared using Bayes Factor (BF), available in TRACER program (version 1.4: <http://beast.bio.ed.ac.uk/Tracer>). The uncertainty in the data is reflected in the 95% highest probability density (HPD).

RESULTS

Phylogenetic signal of HCV sequences.

The phylogenetic signal of the complete HCV genome and different genomic regions were analysed with the likelihood mapping method. As shown in Table 2, the full-length genome has the highest phylogenetic information (98.1%). The lowest signal contents were observed in the 5'UTR, p7, and NS4A, ranging from 62.3% to 76.4%. The regions that exhibited the highest phylogenetic information were NS5B (95.8%), NS3 (95.7%) and NS5A (93.3%) respectively.

Table 1. Detailed of HCV sequences included in this study.

Name	Genotype	Accession	Year	Country	Note
H77	1a	AF009606	1997	United States	
PHCV-1/SF9_A	1a	AF271632	2000	United States	chimpanzee
COLONEL	1a	AF290978	2000	United States	chimpanzee
HEC278830*	1a	AJ278830	2000	United Kingdom	
HC-J1	1a	D10749	1992	Japan	
HCV-1 HCV-PT	1a	M62321	1991	United States	chimpanzee
HCV-H*	1a	M67463	1992	United States	
MD1-0	1b	AF165045	1999	Japan	
HCV-S1	1b	AF356827	2001	Singapore	
HCV-TR1*	1b	AF483269	2002	Turkey	
HCV-A	1b	AJ000009	1997	Australia	
HC-C2	1b	D10934	1993	China	
N589*	1b	AY587844	1999	Russian Federation	
J33	1b	D14484	1991	Japan	
HCV-JS	1b	D85516	1996	Japan	
D89815	1b	D89815	1996	Japan	
HCV-J*	1b	D90208	1990	Japan	
HEBEI	1b	L02836	1992	China	
HPCGENANTI	1b	M84754	1992	Taiwan	
HCV-L2	1b	U01214	1993	South Korea	
HCU16362	1b	U16362	1994	South Korea	
HD-1	1b	U45476	1996	Germany	
AY051292	1c	AY051292	2001	India	
Khaja1	1c	AY651061	2004	India	
HC-G9*	1c	D14853	1993	Indonesia	
AJ851228*	1	AJ851228	2004	Equatorial Guinea	
NDM228	2a	AF169002	1999	Japan	
G2AK3	2a	AF169004	1999	Japan	
NDM59	2a	AF169005	1999	Japan	
JFH-1	2a	AB047639	2000	Japan	
AY746460*	2a	AY746460	2004	Japan	
HC-J6*	2a	D00944	1991	Japan	
JPUT971017	2b	AB030907	1999	Japan	
MD2b1-1*	2b	AY232730	2003	Japan	
HC-J8*	2b	D10988	1992	Japan	
BEBE1*	2c	D50409	1995	Italy	
VAT96*	2k	AB031663	1999	Moldova	
D54*	2i	DQ155561	2002	Vietnam	
CB	3a	AF046866	1998	Australia	
NZL1*	3a	D17763	1993	Japan	
K3A	3a	D28917	1994	United States	
HCV-CENS1	3a	X76918	1995	Germany	
HCV-Tr*	3b	D49374	1994	Japan	
JK049*	3k	D63821	1995	Indonesia	
01-09	4a	DQ418782	2006	Egypt	
F753	4a	DQ418787	2006	Egypt	
ED43*	4a	Y11604	1997	Egypt	
03-18*	4d	DQ418786	2006	Egypt	
24	4d	DQ516083	2006	Egypt	

Table 1. (Continue)

Name	Genotype	Accession	Year	Country	Note
24	4d	DQ516083	2006	Egypt	
IFBT88*	4f	EF589161	2007	France	
SA13*	5a	AF064490	1992	South Africa	
EUH1480	5a	Y13184	1997	United Kingdom	
6a77	6a	DQ480512	2006	Hong Kong	
EUHK2*	6a	Y12083	1997	Hong Kong	
D9	6a	EU246930	2002	Vietnam	
TH580*	6b	D84262	1994	Thailand	
Th846*	6c	EF424629	1999	Thailand	
VN235*	6d	D84263	1996	Vietnam	
GX004	6e	DQ314805	2005	China	
D42*	6e	EU246931	2002	Vietnam	
D88	6e	EU246932	2002	Vietnam	
C-0044	6f	DQ835760	2006	Thailand	
C-0046	6f	DQ835764	2006	Thailand	
TH52*	6f	EU246936	2002	Thailand	
JK046*	6g	D63822	1995	Indonesia	
HK6554	6g	DQ314806	2005	Hong Kong	
VN004*	6h	D84265	1996	Vietnam	
C-0159	6i	DQ835762	2006	Thailand	
Th602*	6i	DQ835770	1994	Thailand	
TH24	6i	EU246935	2002	Thailand	
C-0667	6j	DQ835761	2006	Thailand	
Th553*	6j	DQ835769	1999	Thailand	
KM45	6k	AY878650	2002	China	
KM41	6k	AY878651	2002	China	
VN405*	6k	D84264	1996	Vietnam	
537796	6l	EF424628	2007	United States	
D33*	6l	EU246933	2002	Vietnam	
C-0208	6m	DQ835763	2006	Thailand	
C-0185	6m	DQ835765	2006	Thailand	
C-0192	6m	DQ835766	2006	Thailand	
B4/92*	6m	DQ835767	1995	Thailand	
D86/93*	6n	DQ835768	1994	Thailand	
KM42	6n	AY878652	2002	China	
TH22	6n	EU246937	2002	Thailand	
TH31	6n	EU246938	2002	Thailand	
QC227	6o	EF424627	2005	Canada	Asian immigrant
D85*	6o	EU246934	2002	Vietnam	
QC216*	6p	EF424626	2005	Canada	Asian immigrant
QC99*	6q	EF424625	2004	Canada	Asian immigrant
D49*	6t	EU246939	2002	Vietnam	
TV241	6t	EF632069	2007	Vietnam	
TV249	6t	EF632070	2007	Vietnam	
VT21	6t	EF632071	2007	Vietnam	
D83*	6u	EU246940	2002	Vietnam	
GZ52557*	6	DQ278892	2004	China	
QC69*	7a	EF108306	2003	Canada	Asian immigrant

* samples collected for evolutionary rate and molecular clock analysis

Table 2. Likelihood mapping analysis of the 10,000 possible quartets of the full-genome HCV genotype 1-6 and only genotype 6.

Gene region	No. sites	Constant sites (%)	Tree-like Phylogeny	Star-like Phylogeny
			(phylogenetic signal) ^a Genotype 1-6	(phylogenetic noise) ^a Genotype 1-6
FG^a	9357	30.1	98.1	1.9
5'UTR	237	75.1	62.3	37.7
CORE	573	44.0	91.5	8.5
E1	579	21.4	89.8	10.2
E2	1116	26.6	89.3	10.7
p7	189	15.9	74.5	25.5
NS2	651	20.9	89.3	10.7
NS3	1893	35.6	95.7	4.3
NS4A	162	30.9	76.4	23.6
NS4B	783	30.3	90.9	9.1
NS5A	1434	19.8	93.3	6.7
NS5B	1737	31.8	95.8	4.2

^a FG: Full genome

Best-fit nucleotide substitution model selection and phylogenetic reconstruction.

In order to determine the best model of evolution, we compared the likelihood scores estimated by the ModelTest from either full-length genome or from individual genomic regions using the hLRTs and the AIC method. Different data sets resulted in different best-fit models of nucleotide substitution (Table 3). The commonly used K80 model of evolution was only selected for the 5'UTR by the hLRT method. The models of nucleotide substitution increased in complexity, i.e becoming more parameter rich, for the more diverse regions. All the models incorporated invariant sites (I) and assumed rate heterogeneity (Γ) across sites.

Phylogenetic trees were reconstructed for the full-length HCV genome, NS3, NS5A, and NS5B genes using the GTR + Γ substitution model. Phylogenetic analysis of the full-length genome suggest that genotypes 1 and 4, genotypes 3 and 5, and genotypes 2 and 7 respectively, shared a common ancestor (Fig 1A). In contrast genotype 6 formed a single clade, which based on the branch length, separated earlier than the other genotypes. Furthermore, genotype 2 had a longer evolution time compared to the other genotypes, while genotype 7 split after genotype 1, 3, 4 and 5. The tree based on NS3 (1893 bps, Fig.1B) and NS5A (1434 bps, Fig.1C), had a clustering pattern similar to full-length genome. Branch lengths of the NS3 tree indicated an evolution rate for genotype and subtype similar to the evolution rate of the whole genome. The NS5A tree had

shorter branch lengths, indicating a different evolution rate (Fig.1C). In contrast, analysis of the NS5B (1737 bps, Fig.1D) tree demonstrated a different phylogenetic relationship between the HCV genotypes. The best tree constructed from NS5B sequences shows that genotype 2, 5 and 7 split from genotype 1, while genotype 3 and 4 shared a common ancestor. Genotype 6 is a separate clade.

Table 3. Likelihood values of different nucleotide substitution models for full-length HCV and individual genomic regions.

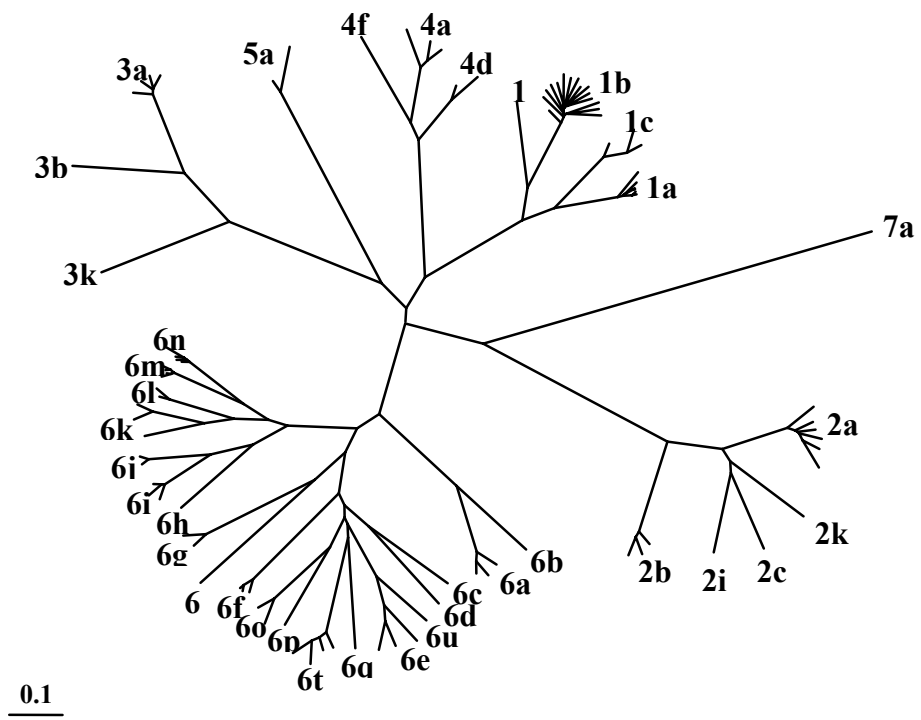
Regions	hLRTs		AICs	
	Best-fit Model	- ln Likelihood	Best-fit Model	- ln Likelihood
FG^a	TrN+I+Γ	311970.5625	GTR+I+Γ	311310.4062
5'UTR	K80+I+Γ	1076.6461	GTR+I+Γ	1045.5182
Core	TrN+I+Γ	11152.3809	TVM+I+Γ	11122.0908
E1	GTR+I+Γ	21913.9805	TVM+I+Γ	21914.0566
E2	GTR+I+Γ	46293.6641	GTR+I+Γ	46293.6641
P7	HKY+I+Γ	7700.5259	GTR+I+Γ	7678.0122
NS2	TrN+I+Γ	26788.3438	GTR+I+Γ	26721.1621
NS3	TrN+I+Γ	59567.1328	GTR+I+Γ	59458.5625
NS4a	GTR+I+Γ	5063.7207	GTR+I+Γ	5063.7207
NS4b	TrN+I+Γ	25488.7715	GTR+I+Γ	25443.5293
NS5a	TrN+I+Γ	51390.1094	GTR+I+Γ	51226.6211
NS5b	TrN+I+Γ	51677.6328	GTR+I+Γ	51559.8750

^a FG: Full genome

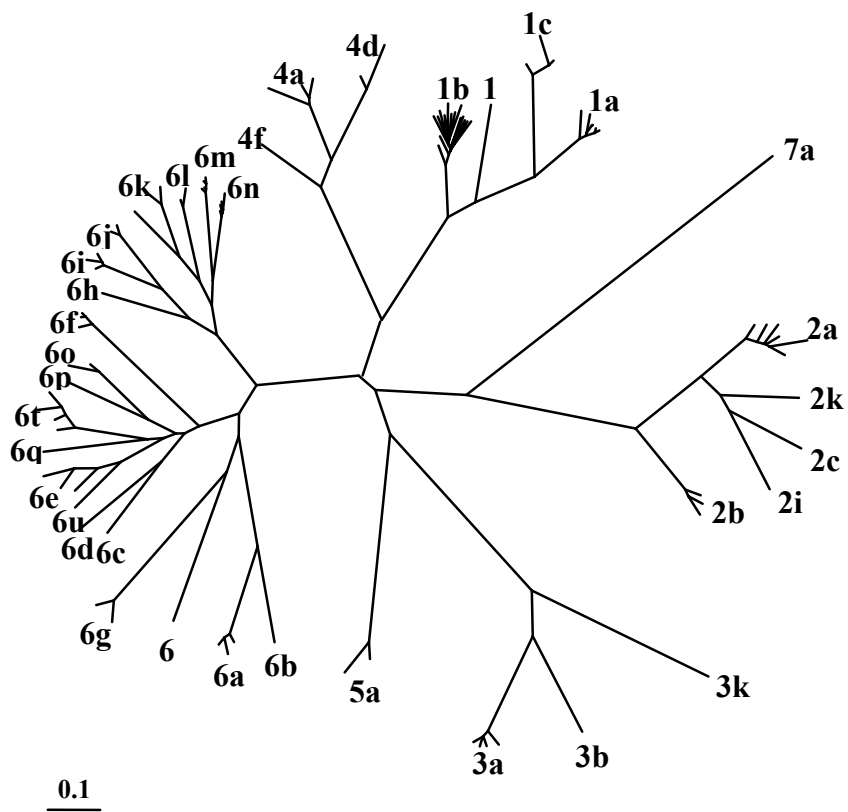
Substitution rates, divergence times, and population dynamics

Since NS3 and NS5A genes were superior representatives of full-length HCV genome evolution, they were used as targets for further analyses. Rates of nucleotide substitution for NS3 was first analysed from 42 sequences using a Bayesian coalescent approach (Table 4). For the NS3 gene, the best-fit population dynamic model for HCV evolution was estimated to be a constant population growth. Strict and relaxed uncorrelated lognormal models were analysed and BF was used to select the best model. The 3-codon partitions (CP3) of rate heterogeneity among sites (GTR+Γ+CP3) showed a substantially superior fit to data than the GTR+I+Γ (Table 4), also when compared to the CP2 without rate heterogeneity gamma distribution (log BF=847.90, data not shown). The substitution rate in NS3 for the CP2 ranged from 6.60 to 13.09 x 10⁻⁴ substitutions/site/year (95% HPD: 0.012-30.09) depending on molecular clock model tested (Table 4).

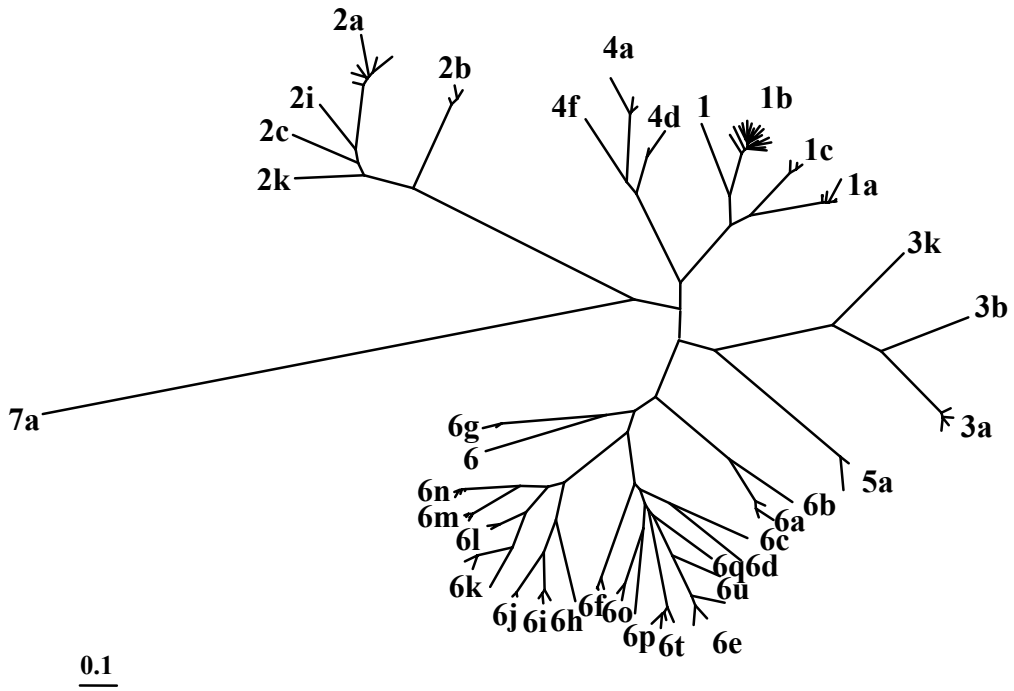
1a)



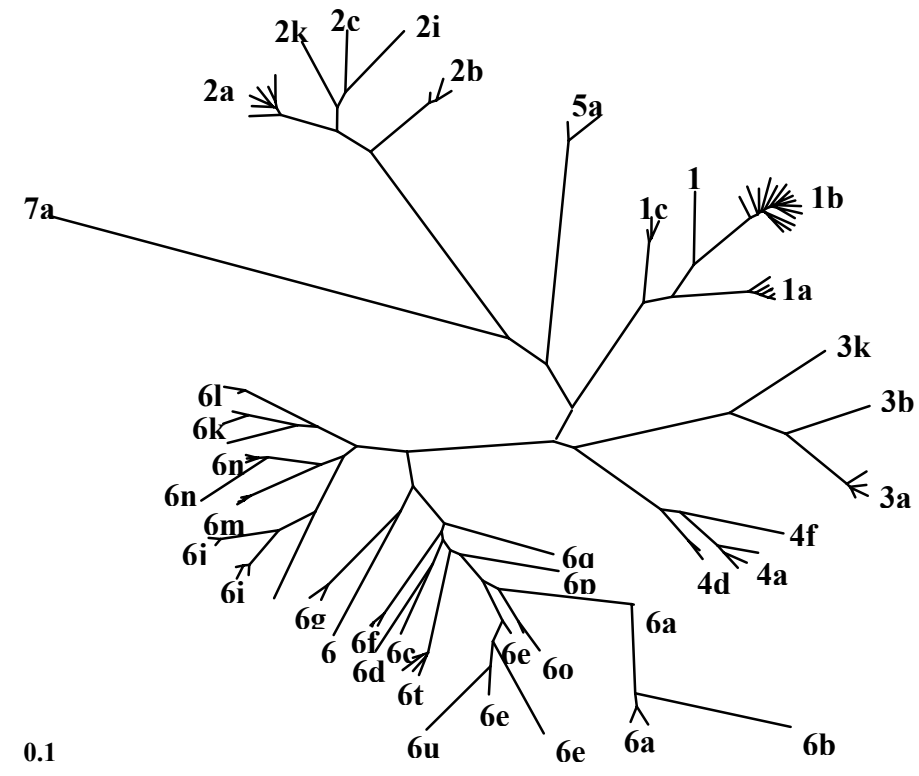
1b)



1c)



1d)



Based on the GTR+ Γ +CP3 model, the mean rate of nucleotide substitution in HCV was then compared for full-length genome, NS3 and NS5A genes. As shown in table 5, similar substitution rates were estimated for NS3 and NS5A genes, at between 14.5 and 18.8×10^{-4} substitutions/site/year and with similar HPD values (Table 5). The mean substitution rate of all genotypes was much smaller across the whole genome (5.71×10^{-4} substitutions/site/year), with HPD values overlapping NS3 and NS5A. This effect is probably due to the presence of high conserved regions such as 5' and 3' UTRs in the full-length genome. The lower 95% HPD value, which represents the lowest possible (most conservative) rates of HCV evolution were 4.4×10^{-6} sub/site/year for full-length genome, 1.6×10^{-6} sub/site/year for NS3 and 7.7×10^{-6} for NS5A gene respectively.

The estimated age of genetic diversity in the NS3 and NS5A genes was broadly similar between the different genotypes (mean estimates of 2639-3893 and 1114-2599 years respectively). Although the estimated age of genetic diversity for the NS5A gene was smaller, the HPD values overlap with those estimated for NS3. In all cases, the strict molecular clock was rejected as a valid description of the evolutionary dynamics of HCV in favour of a relaxed uncorrelated lognormal molecular clock. It is worth noting that the Bayesian Factor (BF) of strict and relaxed clock analysis was less than 20, indicating that the difference between both models was small. The approximated age of the entire tree calculated based on full length genomes was estimated to be about 4610 years (Table 5). Based on the tMRCA estimation for full-length HCV genome, genotype 2 and 7, and genotype 3 and 5 evolved in a period of about 4000-4400 years, while genotype 1 and 4 emerged in the same period as genotype 6, in the range of 2600-3000 years ago. These results correlated with the phylogenetic relationships observed from the branch lengths of genotype 2 and 7 and of genotype 3 and 5 (Fig. 1A).

Figure 1. The best unrooted ML trees depicting the phylogenetic relationship among 96 HCV strains obtained from HCV database. Trees were constructed for a complete genome (A), a complete NS3 codons (B), a complete NS5A codons (C) and a complete NS5B codons (D) datasets using Heuristic ML tree search parameter with Branch-swapping Subtree-pruning-regrafting (SPR) algorithm conducted by PAUP (see Material and methods). Branch lengths are drawn to scale. Boldface numbers along the branches indicates the six major lineages of HCV; subtypes of the viruses are also given (according the new nomenclature). Genbank accession number of all sequences shown in Table 1.

Table 4 Summary of the likelihood scores, substitutions rates and estimated time of different molecular clock model analysis of HCV NS3 gene (42 taxa, 1893 bps)

Substitution Model	Molecular clock model ^a	-ln Likelihood ^b	Substitution rate x10 ⁻⁴ Subs/site/year (95%HPD) ^c	Age, yrs (95% HPD) ^d
GTR + CP2	Strict	40700	6.60 (0.012-14.99)	3766.06 (277.22-10820)
GTR + CP2	Relaxed Uncor. log.	40680 (BF=15.83)	9.15 (0.013-19.67)	3018.24 (211.58-5788.79)
GTR + I + T	Strict	40120	9.47 (0.026-19.32)	2841.09 (233.86-6806.75)
GTR + I + T	Relaxed Uncor. log.	40100 (BF=6.54)	10.10 (0.015-22.82)	3189.93 (221.64-6803.45)
GTR + T + CP2	Strict	38940	12.00 (0.007-25.39)	6363.28 (408.22-11540)
GTR + T + CP2	Relaxed Uncor. log.	38920 (BF=5.44)	13.09 (0.019-30.09)	4765.49 (300.02-12230)
GTR + T + CP3	Strict	38750	14.06 (0.483-28.16)	2311.87 (364.58-6902.87)
GTR + T + CP3	Relaxed Uncor. log.	38740 (BF=4.02)	14.77 (0.013-30.81)	4783.60 (311.58-10880)

^a Relaxed Uncorrelated lognormal^b Bayes Factor , comparison of strict and relaxed molecular clock of each model
^{c,d} (lower 95%HPD - upper 95%HPD)

Additional analysis on dating of HCV has been done using the NS3 and NS5A regions. Molecular clock analysis based on NS3 revealed that genotype 2 and 7 and genotype 3 and 5 are the oldest groups, separated about 3700-3900 years ago, while branches of genotype 1, 4, and 6 separated much later (around 2300-2600 years ago) (Table 5). The analysis based on the NS5A gene showed that estimated evolutionary rate was about 18.8×10^{-4} sub/site/yrs, suggesting a faster evolution than entire genomes and NS3 gene. The mean tMRCA for all genotypes based on NS3 was similar to the entire genome, suggesting that NS3 is a better representative of the whole genome than the NS5A gene.

Table 5 Substitution Rates and dating of HCV full-length genome (8,922 bps), NS3 (1,893 bps), and NS5A gene (1,284 bps) under GTR + Γ + CP3 substitution model and the relaxed molecular clock analysis

Data sets Subs/site/year (95%HPD) ^b	Substitution rate $\times 10^{-4}$	tMRCA ^c (95% HPD)
HCV FG ^a , for all genotypes	5.71 (0.044-12.6)	4610.42 (493.71-15420)
Genotype 1 and 4 (n=10)		3168.00 (343.91-10590)
Genotype 2 and 7 (n=8)		4380.11 (433.90-14730)
Genotype 3 and 5 (n=4)		3917.41 (457.29-13100)
Genotype 6 (n=20)		2596.71 (302.45-8715.57)
HCV NS3 gene, for all genotypes	14.50 (0.016-30.4)	4402.99 (311.63-10360)
Genotype 1 and 4 (n=10)		2639.36 (182.11-6102.68)
Genotype 2 and 7 (n=8)		3748.69 (248.02-8876.83)
Genotype genotype 3 and 5 (n=4)		3893.54 (283.06-9245.59)
Genotype genotype 6 (n=20)		2341.13 (157.90-5568.35)
HCV NS5A gene for all genotypes	18.8 (0.077-37.1)	2599.67 (331.41-6430.00)
Genotype 1 and 4 (n=10)		1114.57 (154.95-2745.19)
Genotype 2 and 7 (n=8)		2599.52 (331.41-6430.00)
Genotype 3 and 5 (n=4)		1355.52 (186.79-3348.76)
Genotype 6 (n=20)		1142.26 (160.95-2846.22)

^a FG: Full genome

^b (lower 95%HPD - upper 95%HPD)

^c Mean age in years

DISCUSSION

It is known that HCV is a fast-evolving virus with an average evolutionary rate similar to human immunodeficiency virus (HIV-1), in the order of 10^{-3} /site/year (11, 33, 37). Because of its high mutation rate, HCV has the characteristic sequence heterogeneity resulting in quasispecies, subtypes and genotypes. In this study, molecular phylogenetic analyses were used to study HCV evolution and divergence times.

Although phylogenetic analysis has been used to clarify the origin and the evolution of HCV (21, 25, 33), some aspects of the evolutionary dynamics of the virus still remain unclear. An important aspect in studying molecular evolution of HCV is obtaining an accurate rate of nucleotide substitution, from which divergence times can be estimated. We have used full-length genome sequences of HCV to evaluate the presence of phylogenetic signal. It seems intuitive that a simple model like K-80 may not adequately represent the complexity of the nucleotide substitution process in RNA viruses such as HIV or HCV (25). We found that full-length genome and the NS5B, NS3 and NS5A genomic regions contain high phylogenetic signal, indicating their power to clarify the relationships between different genotypes and subtypes. This finding is in agreement with findings described by Salemi et al. (25). Furthermore, the complex model GTR+I+ Γ was selected as the best-fit model of nucleotide substitutions for full-length genome, NS3, NS5A, and NS5B data sets. The strong rate heterogeneity across sites in the different HVC genes (not shown) may be explained by the different selective pressures along the genome. It is noteworthy that NS5B, together with CORE and E1 are the recommended regions to use for phylogenetic analysis of HCV strains when the full-length genome is not available (31). However, comparison of trees based on NS3 and NS5A revealed that these genomic regions are more appropriate than NS5B for investigating the phylogenetic relationships between subtypes. The low resolution of NS5B for revealing accurate phylogenetic relationships may be caused by the use of the whole gene sequence. As previously reported, the best region for phylogenetic analysis of HCV is located at nucleotides 8283 to 8624 (333 bps) of NS5B (32). This study raises the awareness that using NS5B sequences for genotyping should be restricted to that specific region, especially for identification of new genotypes or subtypes.

Several groups have estimated the evolutionary rate of HCV (11, 21, 33). In our study, the estimated mean rate of nucleotide substitution across sites of all genotypes was about 5.71×10^{-4} sub/site/year, slightly low compare to the approximated rate observed in long term chronically infected chimpanzee (1.44×10^{-3} sub/site/year) (21). Next, we inferred the tMRCA of HCV genotypes using the Bayesian coalescent approach. In all cases the strict molecular clock was rejected, confirming a previous study describing rejection of the strict molecular clock even when the third-codon position was analysed separately (25). Rejection of the strict molecular clock may be explained by the unequal evolutionary rates among different clades of HCV, which is plausible in light of the different selective pressures exerted by the immune system. A substitution model incorporating the third-codon partitions, and assuming a relaxed uncorrelated lognormal molecular clock, proved superior to estimate rates in the NS3 gene. This is in agreement with the Hillis and Huelsenbeck (10), which suggested that all three codons are involved in HCV evolution. In addition, few sequences are available for genotypes 5 and 7.

Under the relaxed molecular clock model (uncorrelated lognormal), we estimated the tMRCA of the different genotypes, which ranged between 2600 and 4600 years ago, a deeper time estimate of evolution than reported previously (34, 18, 23). Such a timescale suggests that modern dispersal is responsible for the phylogeographic of HCV. Genotype 1a, 1b and 3a have a worldwide distribution with blood transfusion and sharing of contaminated needles between infected drug users representing the common modes of transmission (23). These strains contain low sequence diversity, as a result of rapid spreading of a few original strains. On the other hand, genotype 2 and 6 circulates in African and Asian countries and constitute highly divergent strains. This suggests a long period of HCV infection in these areas through vertical and horizontal transmission (23, 36). Based on the genotype distributions, it was suggested that HCV has been endemic in sub-Saharan Africa and South-East Asia for a considerable time, and that the occurrence of infection in Western and other non-tropical countries represents a relatively recent emergence of infection in new risk groups for infection (19, 30). It remains unclear what forced the earlier divergence of the six major genotypes of HCV. New genotypes may be generated when viruses spread into the new host populations, resulting in mutants able to escape host immune responses. Furthermore, many factors such as co-infection of HCV/HIV in intravenous drug users or a strong selection pressure as a result of the

antiviral therapy may influence HCV evolutionary rates. Further studies on the molecular evolution of HCV will require precise data sets combining phylogenetic data with epidemiological data.

ACKNOWLEDGMENTS

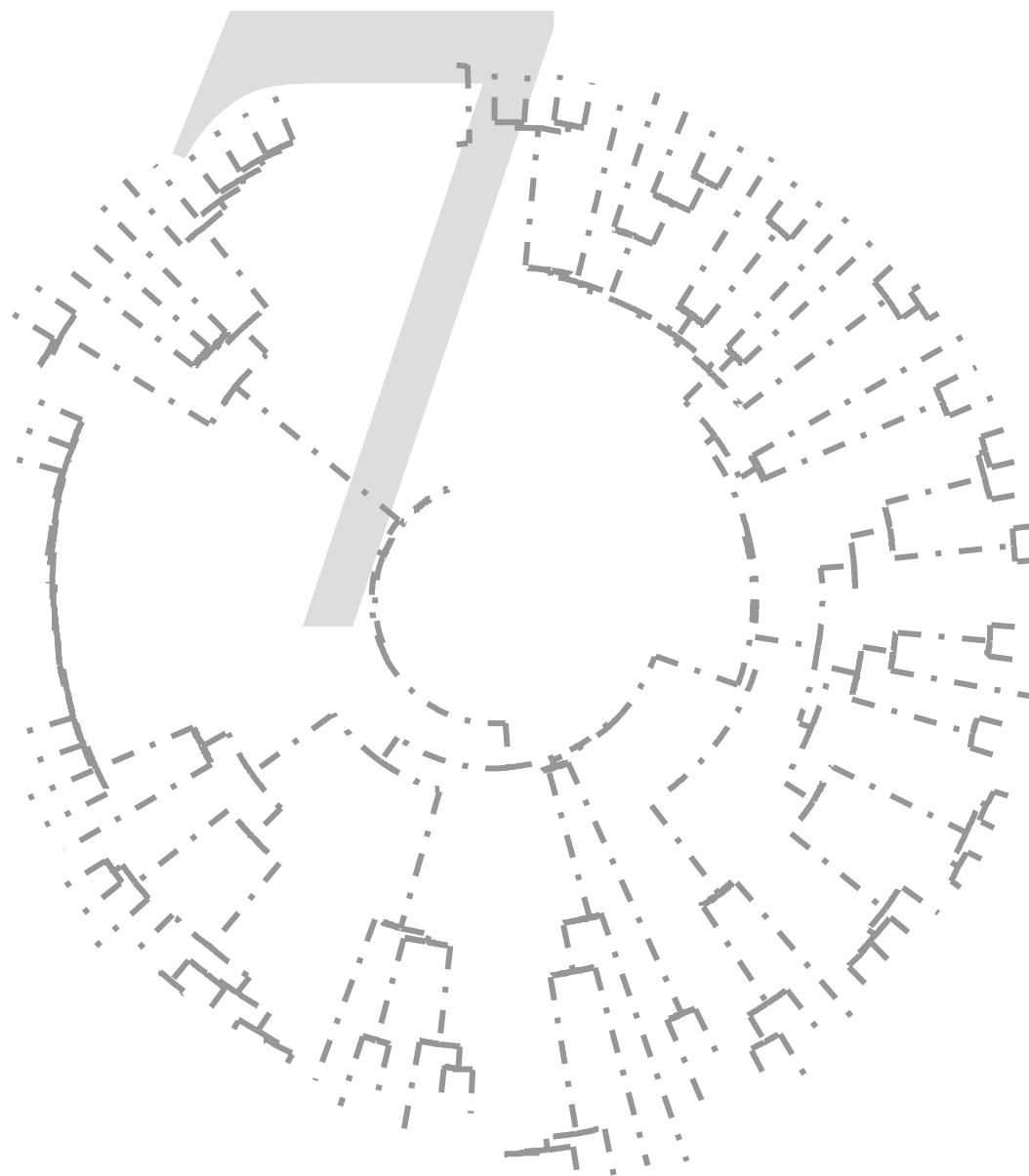
The Commission of the European Community, HECSA Project (ICA4-CT-1999-10009), supported this study. We thank Vu Thuy Yen, Jan Brouwer, Paula van Luijt, Kavita Bedi and Apiradee Theamboonlers for technical assistance.

REFERENCES

1. **Alter, M. J.** 1999. Hepatitis C virus infection in the United States. *J Hepatol* **31 Suppl 1**:88-91.
2. **Booth, J. C., U. Kumar, D. Webster, J. Monjardino, and H. C. Thomas.** 1998. Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. *Hepatology* **27**:223-7.
3. **Bukh, J., R. H. Purcell, and R. H. Miller.** 1994. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proc Natl Acad Sci U S A* **91**:8239-43.
4. **Candotti, D., J. Temple, F. Sarkodie, and J. P. Allain.** 2003. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J Virol* **77**:7914-23.
5. **Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, R. Medina-Selby, P. J. Barr, and et al.** 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A* **88**:2451-5.
6. **Doi, H., C. Apichartpiyakul, K. I. Ohba, M. Mizokami, and H. Hotta.** 1996. Hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand, and identification of novel subtypes of HCV major type 6. *J Clin Microbiol* **34**:569-74.
7. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**:214.
8. **Friebe, P., and R. Bartenschlager.** 2002. Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *J Virol* **76**:5326-38.
9. **Friebe, P., V. Lohmann, N. Krieger, and R. Bartenschlager.** 2001. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J Virol* **75**:12047-57.
10. **Hillis, D. M., and J. P. Huelsenbeck.** 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *J Hered* **83**:189-95.
11. **Ina, Y., M. Mizokami, K. Ohba, and T. Gojobori.** 1994. Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. *J Mol Evol* **38**:50-6.
12. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius.** 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* **76**:4034-43.
13. **Kuiken, C., K. Yusim, L. Boykin, and R. Richardson.** 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics* **21**:379-84.
14. **Makuwa, M., S. Souquiere, P. Telfer, E. Leroy, O. Bourry, P. Rouquet, S. Clifford, E. J. Wickings, P. Roques, and F. Simon.** 2003. Occurrence of hepatitis viruses in wild-born non-human primates: a 3 year (1998-2001) epidemiological survey in Gabon. *J Med Primatol* **32**:307-14.
15. **McHutchison, J. G., S. C. Gordon, E. R. Schiff, M. L. Shiffman, W. M. Lee, V. K. Rustgi, Z. D. Goodman, M. H. Ling, S. Cort, and J. K. Albrecht.** 1998. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *N Engl J Med* **339**:1485-92.

16. **Mellor, J., E. C. Holmes, L. M. Jarvis, P. L. Yap, and P. Simmonds.** 1995. Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J Gen Virol* **76 (Pt 10):**2493-507.
17. **Mellor, J., E. A. Walsh, L. E. Prescott, L. M. Jarvis, F. Davidson, P. L. Yap, and P. Simmonds.** 1996. Survey of type 6 group variants of hepatitis C virus in Southeast Asia by using a core-based genotyping assay. *J Clin Microbiol* **34:**417-23.
18. **Nakano, T., L. Lu, P. Liu, and O. G. Pybus.** 2004. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J Infect Dis* **190:**1098-108.
19. **Ndjomou, J., O. G. Pybus, and B. Matz.** 2003. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *J Gen Virol* **84:**2333-41.
20. **Nixon, K.** 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15:**407-414.
21. **Okamoto, H., M. Kojima, S. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiro.** 1992. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* **190:**894-9.
22. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:**817-8.
23. **Pybus, O. G., M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey.** 2001. The epidemic behavior of the hepatitis C virus. *Science* **292:**2323-5.
24. **Rambaut, A., and L. Bromham.** 1998. Estimating divergence dates from molecular sequences. *Mol Biol Evol* **15:**442-8.
25. **Salemi, M., and A. M. Vandamme.** 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol* **54:**62-70.
26. **Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler.** 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18:**502-4.
27. **Shepard, C. W., L. Finelli, and M. J. Alter.** 2005. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* **5:**558-67.
28. **Shinji, T., Y. Y. Kyaw, K. Gokan, Y. Tanaka, K. Ochi, N. Kusano, T. Mizushima, S. Fujioka, H. Shiraha, A. A. Lwin, Y. Shiratori, M. Mizokami, M. Khin, M. Miyahara, S. Okada, and N. Koide.** 2004. Analysis of HCV genotypes from blood donors shows three new HCV type 6 subgroups exist in Myanmar. *Acta Med Okayama* **58:**135-42.
29. **Sikes, D., and P. Lewis.** 2001. PAUPRat: A tool to implement Parsimony Ratchet searches using PAUP*, beta software version 1.
30. **Simmonds, P.** 2001. Reconstructing the origins of human hepatitis viruses. *Philos Trans R Soc Lond B Biol Sci* **356:**1013-26.
31. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus

- proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-73.
32. **Simmonds, P., D. B. Smith, F. McOmish, P. L. Yap, J. Kolberg, M. S. Urdea, and E. C. Holmes.** 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.
 33. **Smith, D. B., S. Pathirana, F. Davidson, E. Lawlor, J. Power, P. L. Yap, and P. Simmonds.** 1997. The origin of hepatitis C virus genotypes. *J Gen Virol* **78 (Pt 2)**:321-8.
 34. **Smith, D. B., and P. Simmonds.** 1997. Review: molecular epidemiology of hepatitis C virus. *J Gastroenterol Hepatol* **12**:522-7.
 35. **Strimmer, K., and A. von Haeseler.** 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A* **94**:6815-9.
 36. **Stumpf, M. P., and O. G. Pybus.** 2002. Genetic diversity and models of viral evolution for the hepatitis C virus. *FEMS Microbiol Lett* **214**:143-52.
 37. **Suzuki, Y., Y. Yamaguchi-Kabata, and T. Gojobori.** 2000. Nucleotide substitution rates of HIV-1. *AIDS Rev* **2**:39-47.
 38. **Swofford, D.** 2003. PAUP* Phylogenetic analysis using parsimony. Version 4.0 . Sinauer Associates, Sunderland, MA.
 39. **Vos, R. A.** 2003. Accelerated likelihood surface exploration: the likelihood ratchet. *Syst Biol* **52**:368-73.
 40. **WHO.** 1997. Hepatitis C: global prevalence. *Wkly Epidemiol Rec* **72**:341-4.



CHAPTER 7

Detection of Hepatitis C virus specific T cells using Epstein Barr virus based plasmids

Suwanna Noppornpanth¹, Teraporn Chinchai², Sven Bruijns¹, Yong Poovorawan³, Albert D. M. E. Osterhaus¹ and Bart L. Haagmans¹

Department of Virology, Erasmus MC, The Netherlands¹

Department of Microbiology, Faculty of Medicine Srinakharinwirot University, Bangkok, Thailand²

Center of Excellence in Viral Hepatitis Research, Department of Pediatrics, Chulalongkorn Hospital Bangkok, Thailand³

Manuscript in preparation

ABSTRACT

CD8⁺ cytotoxic T lymphocytes (CTLs) play a crucial role in the pathogenesis of hepatitis C virus (HCV) infection. Given the fact that HCV displays enormous sequence diversity and most studies employed genotype 1 based antigens to analyse epitope specificity of CTLs, only limited knowledge exists on genotype cross reactivity of HCV specific CTLs. We therefore explored the possibility to use an Epstein Barr virus (EBV) based plasmid system that allows rapid generation of diverse sets of HCV expressing vectors. First, HLA-A1 restricted HCV specific T cells were obtained from the liver of a chronic HCV patient. These CD8⁺ T cells were shown to specifically recognize recombinant Vaccinia virus (rVV) infected BLCL expressing NS3, when analysed for IFN-gamma production by ELISA, ELISPOT and FACS. Next we cloned different NS3 fragments into EBV based expression plasmids. The efficiency to present HCV antigens by BLCL stably transfected with EBV-based plasmid was similar to BLCL infected with rVV-HCV and enabled the identification of an HLA-A1 restricted epitope. Thus, EBV based expression plasmids can be used as an alternative strategy to detect HCV specific CD8⁺ CTLs.

INTRODUCTION

Cytotoxic T lymphocytes (CTLs) play a major role in the pathogenesis of chronic viral infections (5, 15, 20). Through recognition of virus-infected cells, CTLs may lyse infected cells, or secrete cytokines such as interferon- γ (IFN- γ) and tumor necrosis factor (TNF) that inhibit viral replication and/or recruit other nonspecific inflammatory cells (3, 4, 21). It is likely that cell-mediated immune responses play an important role in viral clearance and/or the clinical outcome of HCV infection. Frequencies of HCV specific CD8⁺ T cells in chronic infected patients, however, are relatively low especially in the peripheral blood (1, 11, 12) and are associated with failure to clear HCV (16).

HCV contains a positive-stranded RNA genome of about 9.4 kilobases, which translates into a single uninterrupted open reading frame that encodes a protein of 3010 amino acids. (2). Sequence diversity from different geographical areas has led to the identification and classification of at least six major genotypes, many of which contain a number of more closely related, yet distinct subtypes (17). The overall sequence similarities of complete genomic sequences is approximately 65-70% between genotypes, and about 80-85% between different subtypes (18). The virus also shows a high degree of variation during the course of infection, resulting in the generation of a number of variants or quasispecies (10).

Assays to monitor HCV specific T cell response require stimulator or target cells presenting viral antigens. Because natural HCV variants cannot infect antigen-presenting cells, alternatives have to be developed to characterise the specificity of the CTL response. A classical way to generate stimulator cells employs autologous Epstein Barr transformed B-cells (BLCLs) infected with recombinant Vaccinia viruses (rVV) encoding parts of the HCV genome (6). However, poxvirus vectors are not convenient to use because of safety issues and the complexity of generating recombinants (8). In this study the use of eukaryotic expression plasmids utilizing CMV or EBV based promoters was considered to transfect BLCL (13, 19). EBV based plasmids may be suitable plasmids for this purposes as these are activated through EBNA-1 from EBV, expressed in BLCLs. (7, 14). We compared the efficiency of EBV-based plasmid and rVVs in expression of HCV-NS3 in BLCLs and evaluated their ability to stimulate HCV specific T cells *in vitro*.

MATERIALS AND METHODS

Patients. Needle liver biopsies and heparinized blood were obtained from chronically HCV-infected patients at the Erasmus MC, Rotterdam, The Netherlands. All patients were infected with HCV genotype 1. Informed consent was written by each patient to participate in our study, and the experiment protocol was approved by the local ethics committee.

Liver-derived T cell lines. Liver biopsies were washed extensively in phosphate-buffer saline (PBS) to remove contaminating blood. To disrupt the hepatic tissue and to release infiltrating lymphomononuclear cells, liver specimens were minced by scalpel and then digested with 0.5 mg/ml collagenase/dipase enzymes and 40 KU/ml *Dnase I* for 30 minutes at 37°C. The suspension was filtrated through a cell strainer (Becton Dickinson) and centrifuged at 1600 rpm for 10 minutes. After discarding the supernatant, the pellet was suspended in tissue culture medium (RPMI-1640) containing 2 mmol/L L-Glutamine, 100 U/ml Penicillin, 100 µg/ml Streptomycin and 10% fetal bovine serum (R10F) in the presence of irradiated allogeneic PBMCs, irradiated BLCLs, 1 µg/ml PHA and 50 IU/ml IL-2. T cell lines were cultured until they stopped proliferating, analysed for CD3, CD56, CD4 and CD8 expression using FACs analysis or frozen for future experiments.

EBV transformed B cells. PBMCs were collected from heparinized blood using Lymphoprep centrifugation at 2,800 rpm for 30 minutes and washed with PBS for 3 times. BLCLs were established through incubation with Epstein-Barr virus (EBV) overnight at 37°C. Transformed B cells were washed with medium and cultured in R10F medium.

Vaccinia virus vectors and peptides. Recombinant vaccinia viruses (rVV) expressing HCV genes derived from the HCV-1 strain (genotype 1a) were kindly provided by Chiron Corporation (Emeryville, CA). The vaccinia constructs used in this study include the vv-Poly encoding aa 1-966 , vv-NS3 encoding aa 364-1619, vv-NS4 encoding aa 1590-2053, vv-NS5A encoding aa 2006-2397 and vv-NS5B encoding aa 2396-3011 of HCV. Wild type vaccinia virus was used as control.

Synthetic peptides (ATDALMTGY and CINGVCWTV) corresponding to HLA-A1 and HLA-A2 restricted epitopes, respectively, were purchased from Eurogentec. The

peptides were selected, based on location in NS3-NS4 region of HCV genome (aa 966-1619) using the SYFPEITHI program (<http://syfpeithi.bmi-heidelberg.com>).

Cloning and expression of NS3 gene. The 1,980 bp fragment located in NS3 gene (aa 966-1,619) was amplified from vv-NS3 construct using forward primer NNRdF (5'-CAGTATGGATCTGGCCGTGGCTGT-3') and reverse primer NNRdR (5'-AGGGTGGGCTTGAGGCGAAT-3'). The amplified PCR product was ligated into pCR2.1 vector (TA cloning ® kit, Invitrogen) and transformed into *E.coli* INVαF' according to the manufacturer's protocol. Positive clones were selected by X-gal/IPTG and ampicillin resistant. Recombinant plasmid was purified and inserted DNA was digested with the restriction enzyme *EcoR* I followed by a fill-in and dephosphorylation steps. DNA was subcloned into a VR1012 Neo⁺ mammalian expression vector (Vical) at the *EcoR* V site and subcloned again into pNS EBV-based vector at the *BamH* I site (Fig. 1). Positive clones were selected by kanamycin resistance and plasmid DNA were extracted and examined by restriction enzyme digestion. Recombinant DNA was purified for the transfection experiment using the StrataPrep EF plasmid Midiprep Kit (Stratagene) according to the manufacturer's protocol. To specify the specific region recognized by specific T cells, small fragments containing different parts of the NS3 gene were amplified and subcloned into the pNS vector. Positive clones of construct 4 (aa. 966-1135), construct 5 (aa. 1126-1344) and construct 6 (aa. 1328-1617) were selected and purified as previously described.

Purified pNS-NS3 and control vectors were transfected into NKNT-3 cells by using Lipofectamine 2000 reagent (Life Technologies) or into BLCLs by electroporation using a Gene Pulser apparatus (BioRad Laboratories) pulsed at 390 V/500µF. Cells were selected for stable transfection by adding 1 mg/ml G418 sulfate in RPMI 1640 medium on day 3 after transfection. Transfected NKNT-3 cells were cultured for 24-48 hrs and fixed with 5% paraformaldehyde in phosphate-buffered saline (PBS). After washing with PBS, cells were incubated with 1.0 M glycine for 10 minutes followed by 0.05% H₂O₂ in 70% ethanol. Non-specific binding was blocked with 1% normal goat serum. The cells were incubated with a mouse monoclonal antibody to NS3 (5F-1, Organon) at 4° C for 12 hrs. After incubating with biotin conjugated goat anti-mouse IgG and streptavidine peroxidase reagent, substrate was added (3'-amino-9-ethylcarbazole, AEC) in 50 mM

acetate buffer pH5.0 and examined under light microscopy. BLCL were stained similarly except that a goat anti-mouse-FITC antibody was used as a second step.

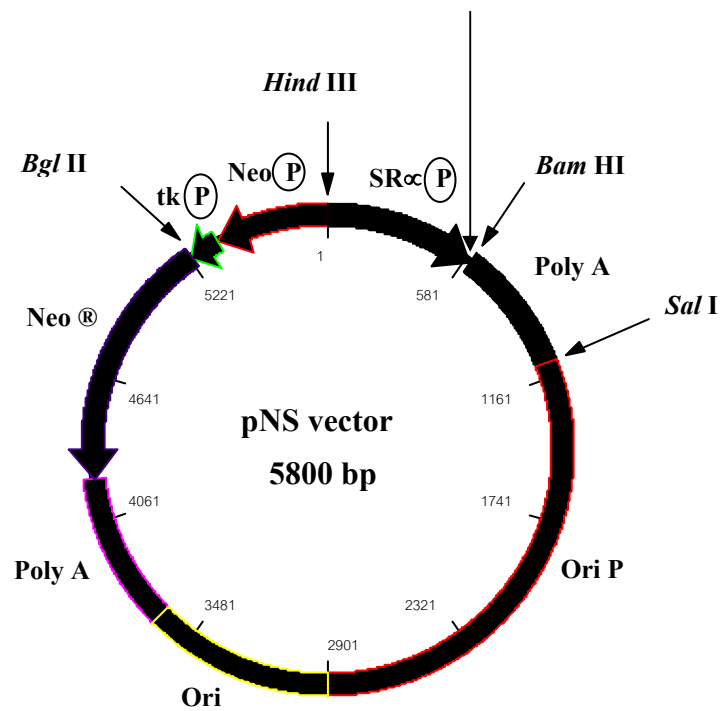
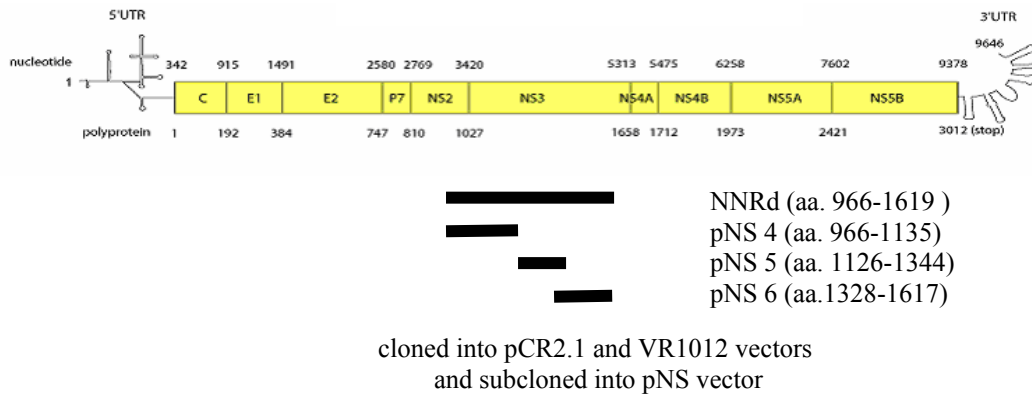


Figure 1 clone strategy of HCV NS3 genes amplified from rVV construct (genotype 1). Different PCR fragments (NNRd, pNS 4-6) were separately cloned into pCR 2.1 and subcloned into VR1012 and pNS vectors, respectively.

FACS analysis for IFN- γ production. Cells were stimulated for 2 hours with autologous BLCLs infected with recombinant vaccinia-HCV, autologous BLCLs stably transfected with pNS-HCV and control autologous BLCLs. Golgi-Stop was added at a ratio of 4 μ l Golgi Stop and 6 ml medium and further incubated for 4 hours. Cells from each well were washed and suspended in Cytifix/Cytoperm solution for 20 minutes at 4°C. The permeabilized cells were washed 2 times with 1x Perm/Wash solution (Pharmlingen) and stained for intracellular IFN- γ production by using anti-IFN- γ -PE and isotype control for 30 minutes in the dark on ice. After this the cells were washed 2 times with 1x Perm/Wash solution and resuspended in 250 μ l staining buffer before FACS analysis. FACS analysis was performed by FACScan flow cytometers and analyzed with Cell Quest software (Beckton Dickinson, San Jose, CA)

ELISA and ELISPOT assay for IFN- γ production Nitrocellulose-bottom Silent Screen 96-well plates (Nalge Nunc International) were coated with 100 μ l of the IFN- γ monoclonal antibody (Mabtech) at a concentration of 1 μ g/ml in 0.1 M carbonate bicarbonate buffer pH 9.6 and were incubated overnight at 4°C. Unbound antibodies were removed via 3 successive washings with PBS. The coated wells were blocked with culture medium (RPMI-1640 and 10% human serum) for 2 hours in 37°C incubator. Wells were washed and filled with 100 μ l R10F containing 1×10^5 T cells together with 2×10^4 irradiated BLCLs infected with VV expressing HCV or transfected with pNS vectors. PHA (1 μ g/ml) or PMA was used as a positive control (these mixtures were previously incubated in round bottom 96 well plate in 37°C incubator for 4 hours before transferred to ELISPOT plate). The samples were incubated for 24 hours at 37°C incubator with 5% CO₂. After incubation, cells were removed by washing the plate 6 times with PBS containing 0.05% Tween 20. The biotin-conjugated anti-IFN- γ (Mabtech AB) was added to each well at the concentration of 1 μ g/ml and incubated for 3 hours at room temperature. The plates were rinsed 3 times by immersion in PBS containing 0.05% Tween 20 and were exposed to 100 μ l of streptavidin-alkaline phosphatase (Mabtech AB) for 1 hour. Unbound conjugate was removed by washing with PBS, and finally, 100 μ l of 5-bromo-4-chloro-3-indolylphosphate/nitroblue tetrazolium substrate solution (BCIP/NBT, Bio-Rad Laboratories, Richmond, CA) was added, and the sample

was incubated for 40 minutes. The color reaction was stopped by extensive washings 3 times with distilled water, and after dry, the number of spots was scored by using the dissection microscope. IFN gamma in supernatants was determined by ELISA according the manufacturer's guidelines (U-Cytech).

T cell depletion and HLA restriction analysis. Depletion and isolation of CD4⁺ or CD8⁺ T cells was performed using DYNABEADS M-450 (111.05 and 111.07; Dynal, Oslo, Norway) according to the manufacturer's protocol. Depleted CD4⁺ or CD8⁺ were stimulated with rVV-NS3 and VV wild type. The positive control C1R cell expressing HLA-A1 restrict and wild type were kindly provided by Dr. A.C.M. Boon (Virology department, Erasmus MC, Rotterdam). Both cells line were infected with rVV-NS3 and VV wild type.

RESULTS AND DISCUSSION

Generation of T cell lines from liver biopsies.

We collected liver biopsies and heparinized blood from fifteen chronically HCV infected patients. All patients were infected with HCV genotype 1 and previous non-responder to IFN- α and ribavirin therapy. T cell lines from all patients were analyzed by FACS analysis to determine the percentage of CD4⁺ and CD8⁺ T cells. Most of the CD3⁺ cells expressed the CD8 marker (Fig. 2A and 2B). Next, T cell lines generated from liver biopsies were tested for HCV specificity using recombinant VVs expressing different fragments of the HCV genome. No specific IFN-gamma producing cells were detected with exception from patient 816 (Fig. 2C). This T cell line specifically recognised BLCL infected with VV-NS3. Similar results were obtained using T cell lines from the same patient obtained at two different time points during subsequent IFN- α and ribavirin therapy (not shown). Similarly to our finding, low frequencies of intrahepatic HCV specific T-cells have been observed in other studies examining the presence of intrahepatic HCV specific T-cells in previous non responders to IFN therapy (9).

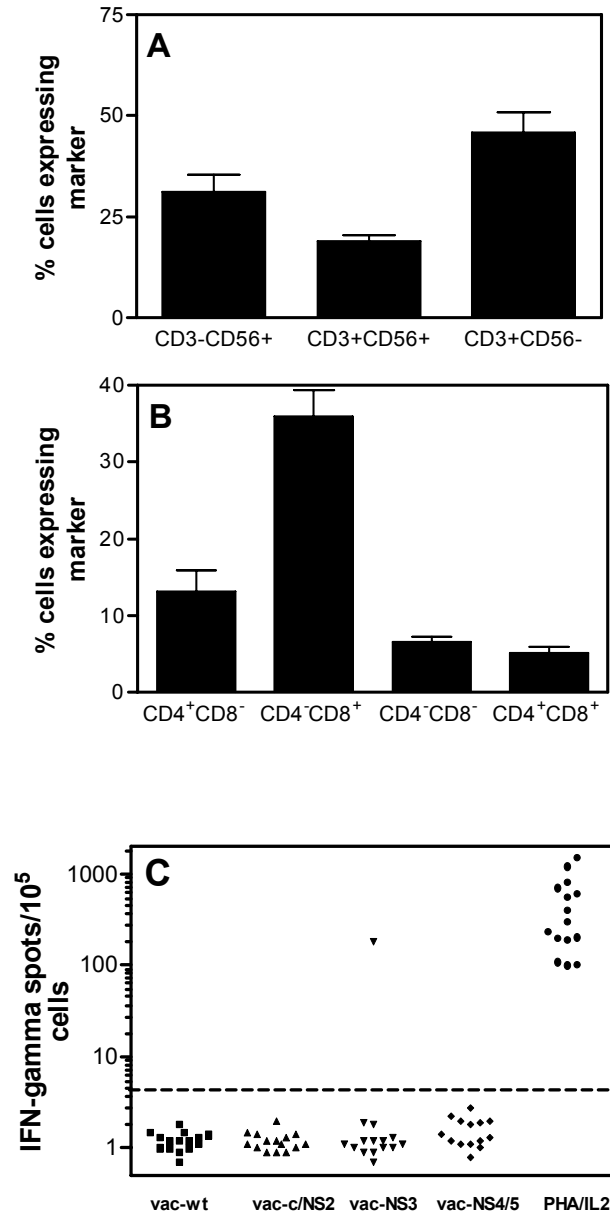


Figure 2. Characterization of intra-hepatic T cell lines from chronic hepatitis C patients. T cell lines from fifteen HCV infected patients were analysed by FACS for cell surface markers CD3 and CD56 (A) and CD3⁺ cells were further characterized for expression of CD4 and CD8 (B). HCV specific responses were analysed for IFN gamma secretion using recombinant VV infected BLCL in an ELISPOT assay (C).

HLA restriction.

The T cell line of patient 816 was further analysed for HLA restriction using T cell depletion and HLA restriction studies. As shown in figure 3A, depletion of CD8⁺ cells from the T-cell line abrogated the specific T cell response whereas no effect was observed when applying CD4 depletion. When a panel of HLA typed BLCL was infected with recombinant VV expressing NS3, only those expressing HLA-A1 were specifically recognized by the T-cells (not shown). To confirm HLA-A1 restriction, C1R cells expressing HLA-A1 were infected with VV expressing NS3 (Figure 3B).

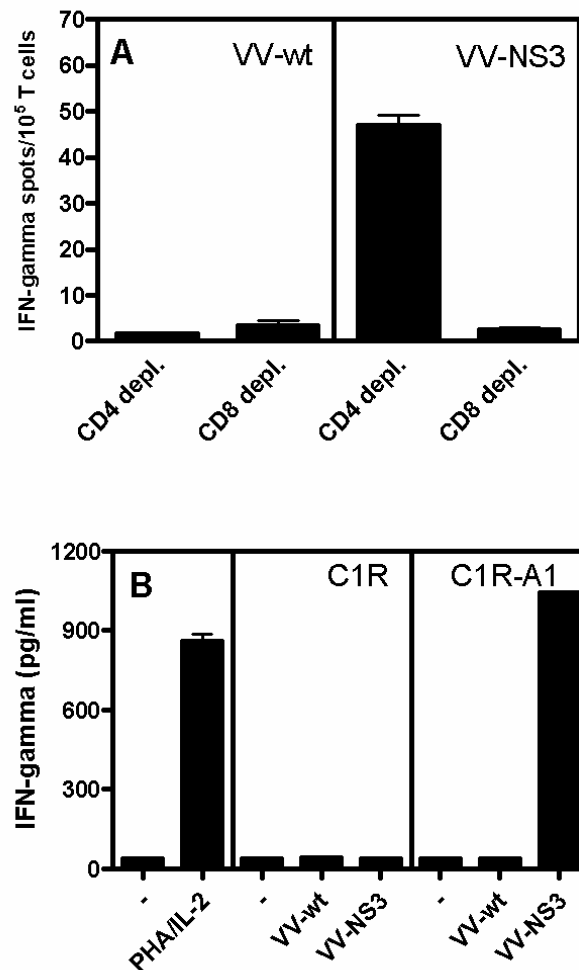


Figure 3. HLA restriction of T cell line 816. Depletion of CD4 and CD8 positive T cells from T cell line 816 (A) and restriction through HLA-A1 as determined in CR1 and CR1 HLA-A1 transfected cells infected with recombinant VV expressing NS3 analysed using IFN-gamma ELISA (B).

pNS expression plasmids encoding NS3.

In order to further characterize the epitope recognized by the T cells, we utilized Epstein Barr based plasmids. The EBV-based plasmids that we used in the present study (pNS vector, Fig.1) carry the oriP region from EBV as *cis* elements for DNA replication. After transfection into BLCLs, EBNA-1 from EBV in BLCLs binds to oriP and facilitates the retention and replication of the plasmid DNA, making pNS vector a very efficient expression vector (7)

First, we amplified the HCV NS3 gene (from rVV-NNRd) using primer NNRdF and NNRdR. The PCR product was cloned into the TA cloning vector and subcloned into VR1012 and pNS vectors as described in materials and methods (Fig.1). The efficiency of these vectors to express NS3 protein was first verified in NKNT-3 cells. Plasmids carrying the CMV promoter transfected into NKNT-3 cells using Lipofectamine 2000 were shown to express NS3 but no staining was observed when pNS-NS3 was transfected, confirming the promoter specificity (not shown). In contrast, when BLCL were electroporated with pNS-NS3 and subsequently selected under Geneticin pressure, we observed significant expression of NS3 (Figure 4). In contrast, relatively low levels of NS3 were detected in VR-NS3 transfected BLCL (not shown).

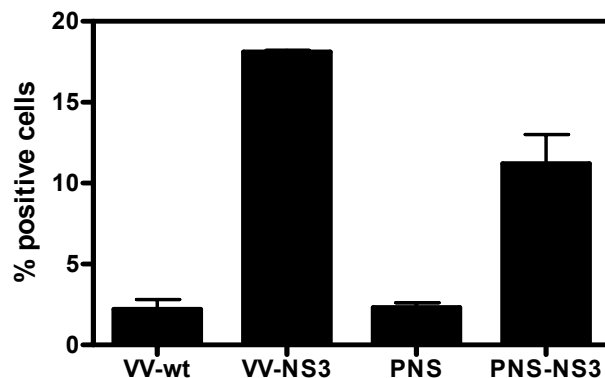


Figure 4. NS3 antigen expression in pNS transfected cells. BLCL infected with wt VV, recombinant VV expressing NS3, or transfected with pNS and pNS-NS3 were stained for NS3 using a monoclonal antibody and subsequent Facs analysis.

Epitope specificity of T cell line 816

Next we tested the antigen presenting properties of pNS transfected cells using three different methods to detect IFN- γ production (ELISPOT, ELISA and intracellular FACS staining). As a positive control, T cells were stimulated with PHA or PMA. As shown in Fig. 5A, T cells specifically produced IFN- γ when incubated with BLCL transfected with pNS-NS3 or BLCL infected with rVV-NS3. The amount of IFN- γ produced was not significantly different using both types of stimulator cells. Similar results were obtained using ELISPOT or intracellular FACS analysis (not shown).

To further pinpoint the region that contains the HLA-A1 restricted epitope, three overlapping fragments of the NS3 gene were cloned into the pNS vector. Only BLCLpNS6.1 that contains aa. 1328-1617 of the NS3 gene was specifically recognized (Fig 5B), enabling the identification of the epitope recognized by the IFN-gamma producing T cells through bioinformatics analysis of potential HLA-A1 restricted epitopes present in the encoded protein. A peptide located in the NS3 coding region of plasmid pNS6.1, ATDALMTGY, was specifically recognized when peptide pulsed BLCL were used to restimulate the T-cell line 816. No IFN-gamma was produced when BLCL were pulsed with HLA-A2 restricted control peptide (also located in NS3). Thus, EBV expression vectors encoding parts of the HCV genome are extremely effective in activating HCV specific CTL when transfected into BLCL. Although autologous BLCL infected with recombinant VVs encoding parts of the HCV genome can be used to monitor HCV specific CTL responses they pose several disadvantages such as infectivity, cytopathic effect and difficulty to generate recombinant vectors. Alternatively, overlapping peptides covering a specific HCV genome may be used to pulse antigen-presenting cells, but this technique remains quite costly when multiple genomes have to be screened. Therefore, EBV based plasmid expression vectors offer an alternative method to monitor HCV specific CTL responses in HCV chronic infected patients.

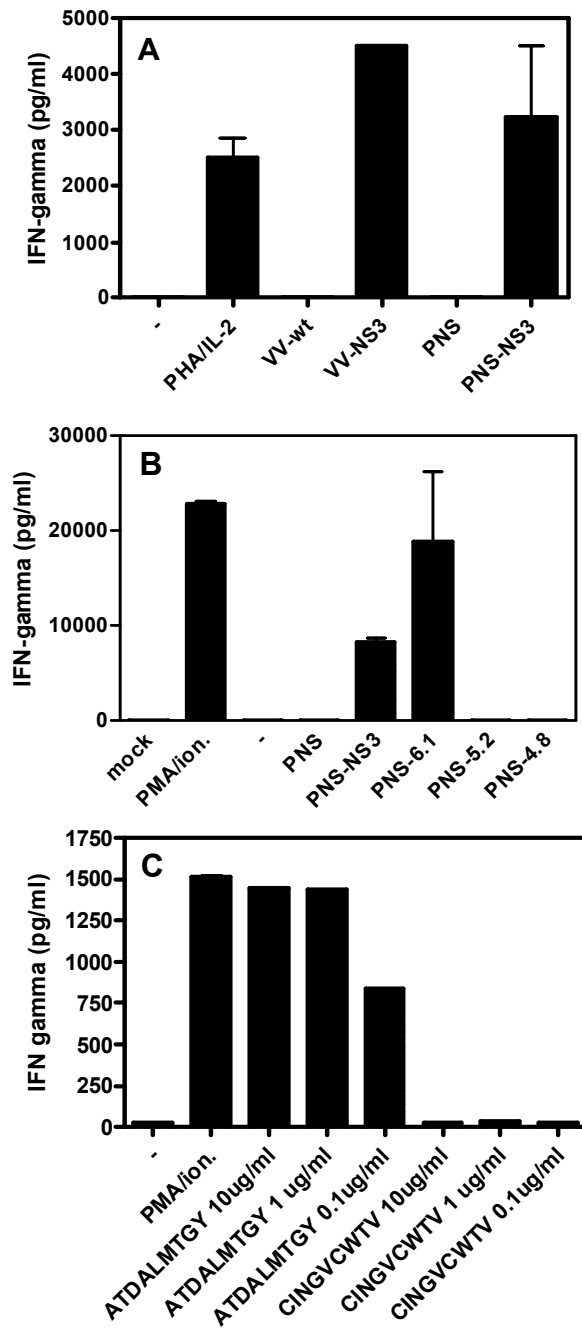
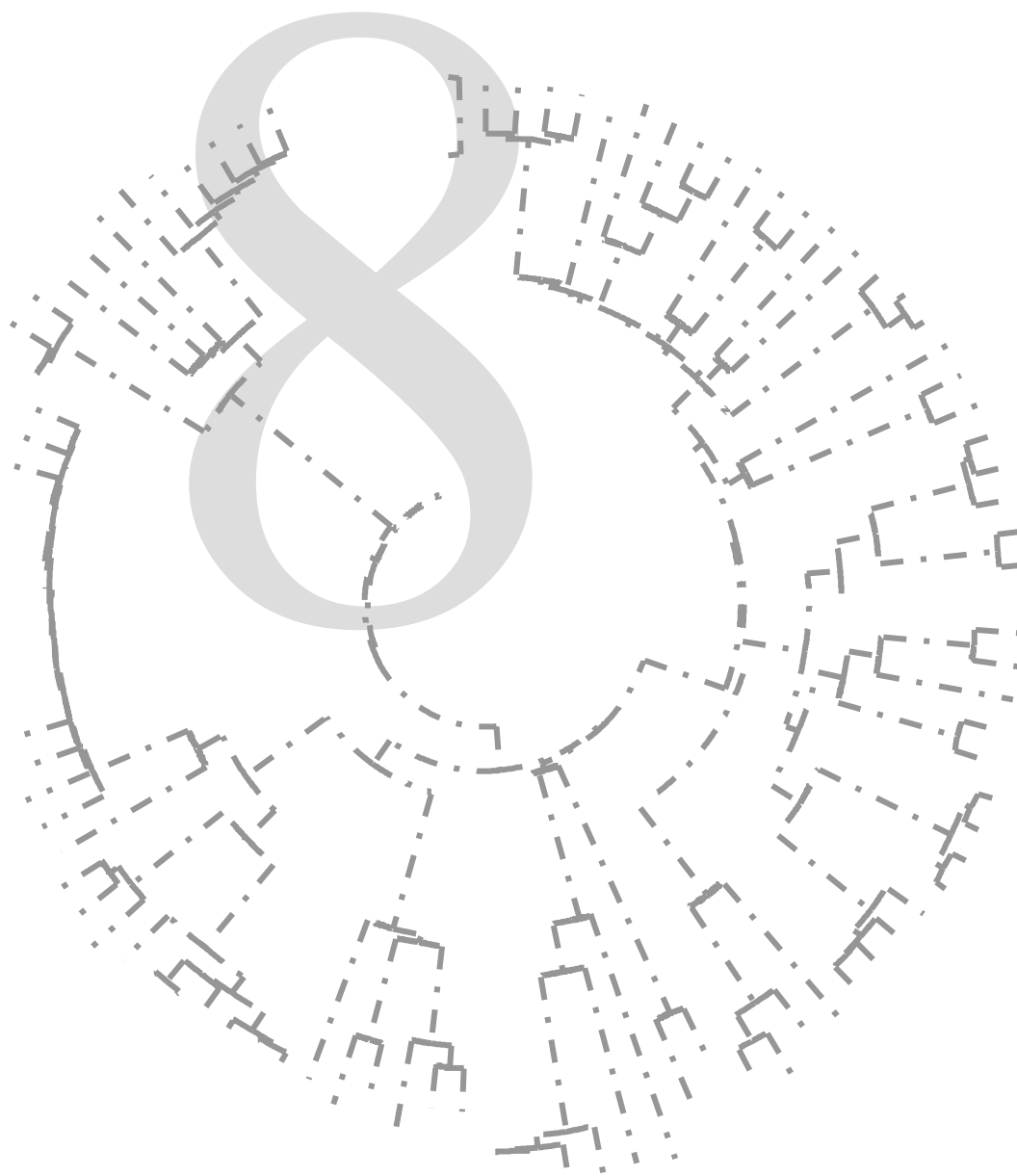


Figure 5. Epitope mapping of T cell line 816 using pNS based expression plasmids. BLCL infected with wt VV, recombinant VV expressing NS3, or transfected with pNS and pNS-NS3 were used to determine IFN gamma production (A). Plasmids expressing overlapping NS3 fragments (pNS6.1, pNS5.2, pNS4.8) were used to further localize the epitope (B), subsequently identified by bioinformatic analysis (C).

REFERENCES

1. **Cerny, A., P. Fowler, M. A. Brothers, M. Houghton, H. J. Schlicht, and F. V. Chisari.** 1995. Induction in vitro of a primary human antiviral cytotoxic T cell response. *Eur J Immunol* **25**:627-30.
2. **Choo, Q. L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**:359-62.
3. **Guidotti, L. G., K. Ando, M. V. Hobbs, T. Ishikawa, L. Runkel, R. D. Schreiber, and F. V. Chisari.** 1994. Cytotoxic T lymphocytes inhibit hepatitis B virus gene expression by a noncytolytic mechanism in transgenic mice. *Proc Natl Acad Sci U S A* **91**:3764-8.
4. **Guidotti, L. G., T. Ishikawa, M. V. Hobbs, B. Matzke, R. Schreiber, and F. V. Chisari.** 1996. Intracellular inactivation of the hepatitis B virus by cytotoxic T lymphocytes. *Immunity* **4**:25-36.
5. **Koelle, D. M., C. M. Posavad, G. R. Barnum, M. L. Johnson, J. M. Frank, and L. Corey.** 1998. Clearance of HSV-2 from recurrent genital lesions correlates with infiltration of HSV-specific cytotoxic T lymphocytes. *J Clin Invest* **101**:1500-8.
6. **Koziel, M. J., D. Dudley, N. Afdhal, A. Grakoui, C. M. Rice, Q. L. Choo, M. Houghton, and B. D. Walker.** 1995. HLA class I-restricted cytotoxic T lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. *J Clin Invest* **96**:2311-21.
7. **Mazda, O., E. Satoh, K. Yasutomi, and J. Imanishi.** 1997. Extremely efficient gene transfection into lympho-hematopoietic cell lines by Epstein-Barr virus-based vectors. *J Immunol Methods* **204**:143-51.
8. **Moss, B.** 1991. Vaccinia virus: a tool for research and vaccine development. *Science* **252**:1662-7.
9. **Nelson, D. R., C. G. Marousis, T. Ohno, G. L. Davis, and J. Y. Lau.** 1998. Intrahepatic hepatitis C virus-specific cytotoxic T lymphocyte activity and response to interferon alfa therapy in chronic hepatitis C. *Hepatology* **28**:225-30.
10. **Qin, H., N. J. Shire, E. D. Keenan, S. D. Rouster, M. E. Eyster, J. J. Goedert, M. J. Koziel, and K. E. Sherman.** 2005. HCV quasispecies evolution: association with progression to end-stage liver disease in hemophiliacs infected with HCV or HCV/HIV. *Blood* **105**:533-41.
11. **Rehermann, B.** 1999. Cellular immune response to the hepatitis C virus. *J Viral Hepat* **6 Suppl 1**:31-5.
12. **Rehermann, B., K. M. Chang, J. G. McHutchison, R. Kokka, M. Houghton, and F. V. Chisari.** 1996. Quantitative analysis of the peripheral blood cytotoxic T lymphocyte response in patients with chronic hepatitis C virus infection. *J Clin Invest* **98**:1432-40.
13. **Retiere, C., V. Prod'homme, B. M. Imbert-Marcille, M. Bonneville, H. Vie, and M. M. Hallet.** 2000. Generation of cytomegalovirus-specific human T-lymphocyte clones by using autologous B-lymphoblastoid cells with stable expression of pp65 or IE1 proteins: a tool to study the fine specificity of the antiviral response. *J Virol* **74**:3948-52.
14. **Robertson, E. S., T. Ooka, and E. D. Kieff.** 1996. Epstein-Barr virus vectors for gene delivery to B lymphocytes. *Proc Natl Acad Sci U S A* **93**:11334-40.

15. **Rooney, C. M., C. A. Smith, C. Y. Ng, S. K. Loftin, J. W. Sixbey, Y. Gan, D. K. Srivastava, L. C. Bowman, R. A. Krance, M. K. Brenner, and H. E. Heslop.** 1998. Infusion of cytotoxic T cells for the prevention and treatment of Epstein-Barr virus-induced lymphoma in allogeneic transplant recipients. *Blood* **92**:1549-55.
16. **Schirren, C. A., M. C. Jung, T. Worzfeld, M. Mamin, G. Baretton, J. T. Gerlach, N. H. Gruener, R. Zachoval, M. Houghton, H. G. Rau, and G. R. Pape.** 2001. Hepatitis C virus-specific CD4⁺ T cell response after liver transplantation occurs early, is multispecific, compartmentalizes to the liver, and does not correlate with recurrent disease. *J Infect Dis* **183**:1187-94.
17. **Simmonds, P.** 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol* **85**:3173-88.
18. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-73.
19. **Teshigawara, K., and Y. Katsura.** 1992. A simple and efficient mammalian gene expression system using an EBV-based vector transfected by electroporation in G2/M phase. *Nucleic Acids Res* **20**:2607.
20. **Walter, E. A., P. D. Greenberg, M. J. Gilbert, R. J. Finch, K. S. Watanabe, E. D. Thomas, and S. R. Riddell.** 1995. Reconstitution of cellular immunity against cytomegalovirus in recipients of allogeneic bone marrow by transfer of T-cell clones from the donor. *N Engl J Med* **333**:1038-44.
21. **Yang, O. O., and B. D. Walker.** 1997. CD8⁺ cells in human immunodeficiency virus type I pathogenesis: cytolytic and noncytolytic inhibition of viral replication. *Adv Immunol* **66**:273-311.



CHAPTER 8

SUMMARIZING DISCUSSION

SUMMARIZING DISCUSSION

HCV infection is a major cause of chronic hepatitis, cirrhosis and hepatocellular carcinoma. Some 170 million persons, translating into 3% of the world population, are chronically infected with HCV, and this number increases annually due to the lack of effective treatments and vaccines (60). A major hallmark of HCV is its genetic variation, which has hampered among others viral classification, diagnosis and vaccine development. In addition, the heterogeneity of HCV plays an important role in the failure of the host's immune response to clear the virus and to protect the host from re-infection or super-infection with different HCV strains. In order to understand the epidemiology and develop effective therapies and vaccines we need to further characterize HCV diversity and the mechanisms that drive the evolution of HCV.

At the start of this study, HCV was classified into 6 genotypes and numerous subtypes. HCV genotypes show different geographical distributions and levels of diversity, reflecting their different epidemic histories and routes of transmission. Sequences reported in the Los Alamos HCV database show that genotypes 1a/b and 3 make up for the majority (~85%) of HCV infections worldwide distributed in Europe, USA, Japan, and Russia (Fig. 1). Genotype 2a infections have been reported from Europe and Japan whereas more divergent genotype 2 subtypes were reported from African countries (5, 16, 35, 44, 59). Genotypes 4 and 5 observed in about 4% of the HCV-infected population, show a high genetic heterogeneity and are mainly restricted to Africa, although cases have been reported from Europe as well (7, 31, 68). To date, genotype 6 viruses were shown to circulate in Thailand, Vietnam, Cambodia, Myanmar, Indonesia, Hong Kong, Singapore, Taiwan and the southern part of China (HCV Los Alamos database). A small number of genotype 6 strains was identified in Europe, Canada, USA and Australia, but mostly these were isolated from immigrants originating from South East Asia (34). The other 2,000 reported HCV sequences (5.45%) are highly diverse (rare subtypes) and several are unassigned (Fig.1). Although prevalent in only 2% of HCV-infected persons, genotype 6 viruses are genetically highly diverse and have been classified in 19 subtypes (6a-6s). Epidemiological data of subtype 6 viruses in South East Asia are limited due to the difficulties related to specimen collection and viral

classification. Therefore, the natural history of HCV genotype 6 infections in this area is unclear.

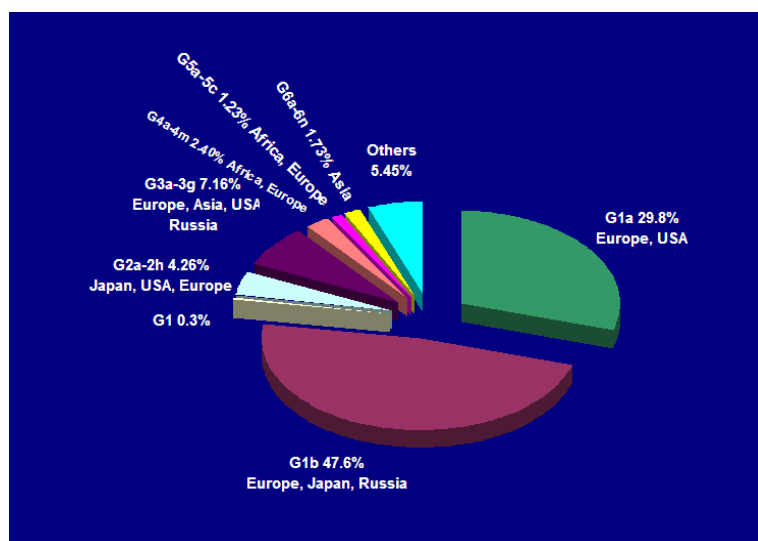


Figure 1. HCV genotype prevalence in the World using sequences deposited in the Los Alamos database.

Most studies on HCV diversity and its role in almost any aspect of HCV biology have been performed on the highly prevalent genotypes 1, 2, and 3. Relatively little is known about whether insights obtained for these genotypes also apply to the less well-studied genotypes, among which genotype 6. In contrast to genotype 1 HCV strains, genotype 6 viruses display more sequence diversity, are basically restricted to a geographical location and appear to be relatively rare, characteristic for endemic transmission (29). Limited clinical symptoms and a relative good response to standard IFN therapy have been reported for genotype 6 viruses (11, 12). In addition, current HCV detection systems designed to detect and classify highly prevalent worldwide strains may fail to detect or classify diverse genotype 6 strains correctly. We studied HCV genotype 6 diversity using serum samples from Thailand and Vietnam (Chapters 2-5) to obtain insights into HCV epidemiology and molecular evolution. Current genotyping/subtyping methods were improved using phylogenetic analysis and a novel line probe assay (chapter 2). Evolutionary mechanisms of HCV genotype 6 were investigated (chapter 3-4), and new subtypes 6 viruses were identified (chapter 5). Bioinformatics tools were applied to study evolution and characterization of HCV

(chapter 6) and molecular tools to detect CTL activity of HCV chronic infected patients are described (Chapter 7).

Genotyping - a novel line probe assay

As a first step towards characterization of the HCV diversity in South East Asia, we analysed HCV-positive samples collected from blood donors in Bangkok Thailand and Ho Chi Minh City, Vietnam in the period from 2000 to 2002. The prevalence of different HCV genotypes was determined using the most practical HCV genotyping test, the VERSANT® HCV genotype assay (INNO-LiPA). This assay is based on the amplification of the conserved 5'UTR region and hybridization to genotype specific probes (53). However, this region contains lower phylogenetic information than other parts of the genome, such as the Core and NS5B genes (45). Not surprisingly, the ability of the VERSANT® HCV genotype assay (INNO-LiPA) to discriminate between HCV genotypes was challenged by the high diversity of HCV variants, especially by the genotype 6c-l variants from SEA that show identical 5'UTR sequences to genotype 1 ((47, 49, 53), chapter 2). The new VERSANT HCV genotype 2.0 assay was developed to increase the accuracy of HCV genotype/subtype identification by including Core specific probes that discriminate between genotype 1a, 1b and 6c-l variants. This assay was used to determine the genotype of HCV viruses in our samples and results were compared to conventional sequence analysis procedures for genotype determination.

As reported in chapter 2, HCV samples from Ho Chi Minh City, Vietnam contained genotype 1 (38%), 2 (10.3%) and 6 (51.7%) sequences, while genotype 1 (40.0%), 3 (45.3%) and 6 (14.7%) were found in the Thai population (chapter 2 and (55)). Overall a relatively high prevalence of genotype 6 viruses was found in both countries. Subtypes characterized from these specimens are 6a, 6d, 6e, 6f, 6i, 6l, 6n and several unassigned or new candidate subtypes. Considering previous reports of HCV diversity in this area, it was not surprising to find a high variation of subtypes 6 viruses since many new subtypes were regularly characterized from Vietnam and Thailand (28, 29, 56, 57).

Molecular diagnostics of HCV is a crucial tool for controlling the epidemic and following the disease progression of HCV. Moreover, clinical studies have shown that the HCV genotype is an important predictor for efficiency of antiviral treatment and

progression of hepatitis (63). Therefore, most treatment protocols require preceding genotype information for HCV infected patients. HCV genotyping relies on the analyses of phylogenetically informative coding regions and comparison to defined genotype sequences. However, sequence analysis of amplicons is laborious and is not successful in identifying mixed infections. Many efforts have been made to develop HCV qualitative and quantitative assays and several commercial tests are available (43). Examination of the HCV positive samples from Thailand and Vietnam by using the old and new VERSANT HCV genotype tests confirmed that a great improvement in accurate genotyping is made with the 5'UTR-Core probe as compared to the 5'UTR probe alone (~96% versus 71%). With regard to the heterogeneity of genotype 6 sequences, 5'UTR-Core line probe assay proved its ability to recognize a broad variety of subtype 6 variants. Moreover, this test also increases the accuracy of genotype 1a/1b identification (4). The new VERSANT[®] HCV Genotype 2.0 assay may actually be the most suitable tool for routine HCV genotyping. This assay may not only be crucial for clinical evaluation of patients, but also for future epidemiological, evolutionary and pathogenesis studies.

HCV recombination

To investigate the accuracy of the newly developed VERSANT HCV genotype 2.0 assay for genotype 6 identification, genotyping of samples was also performed by sequencing Core and NS5B fragments. This effort led us to the identification of an HCV recombinant strain of genotype 2i and 6p (RF2_2i/6p, chapter 3). The recombination break point was identified between nucleotides 3405–3464, which corresponds to the cleavage site of the NS2-NS3 protease that is conserved among different HCV genotypes. Interestingly, the recombination breakpoint of the first identified inter-genotypic recombinant strain of HCV genotype 2k/1b from St. Peterberg, Russia is also located in the NS2/NS3 region of the genome (18). More recently, the 2k/1b recombinant strain was isolated from patients in Ireland (32), Uzbekistan (21) and Estonia (54), suggesting it spreads successfully and is able to compete with wildtype strains.

The fact that both recombinant strains have their recombination breakpoint in the NS2/NS3 region of the genome may suggest that this region is preferable for creating

viable recombinants *in vivo*. Interestingly, the most efficiently propagating HCV chimeras *in vitro* are also recombinants with the breakpoint located in the NS2/NS3 region (26, 40). Currently it is thought that this breakpoint represents the separation point of the genome into two parts, one from NS3-NS5B, involved in and sufficient for replication, whereas the Core-NS2 region encodes structural proteins and proteins needed for assembly of the virus. Depending on the strains fused *in vitro*, viruses with different replication/propagation kinetics arise (40). This may suggest that not all recombinant HCV strains are viable or have an increase in fitness compared to the non-recombinant parental strains *in vivo*.

To investigate whether more recombination events occurred during the diversification of genotype 6 viruses, we collected full-length genomes of 96 HCV from the HCV database, including 35 sequences of genotype 6 viruses. Similarity plot and bootscan analysis identified a new intergenotypic recombination strain, one that was previously identified as subtype 6d in 1998 (57). Strain VN235 (accession number D84263), however, is a recombinant of TH52- (genotype 6f, EU246936) and G2AK3- (genotype 2a, AF169004) like ancestors (Fig. 2). The recombination breakpoints are located in the E1 and E2 genes (position 1468-2289), as was previously observed for the intratypic 1a/1c recombinant (12). These recombination events may have provided the progeny recombinant viruses with a selective advantage by the ability to escape the host immune system by exchanging the E1/E2 region.

Recently, more recombinant forms of inter- and intra-genotype HCV recombinations were described from Vietnam, The Philippines, France and Peru (Table 1), proving our hypothesis that we may be underestimating the true frequency of HCV recombination and that HCV recombination is an important process in the divergence and evolution of HCV (chapter 3). Despite the number of HCV recombinant strains reported nowadays, the true frequency of this event may still be considerably underestimated especially in high endemic areas where multiple genotypes normally coexist in the population. HCV recombinant strains are difficult to detect, especially because most studies use single, short genomic regions to genotype viruses, making the detection of potential recombinants unlikely. Sequence analysis of whole HCV genomes would be an improvement to current genotyping assays and most likely increase the detection of natural HCV recombinants.

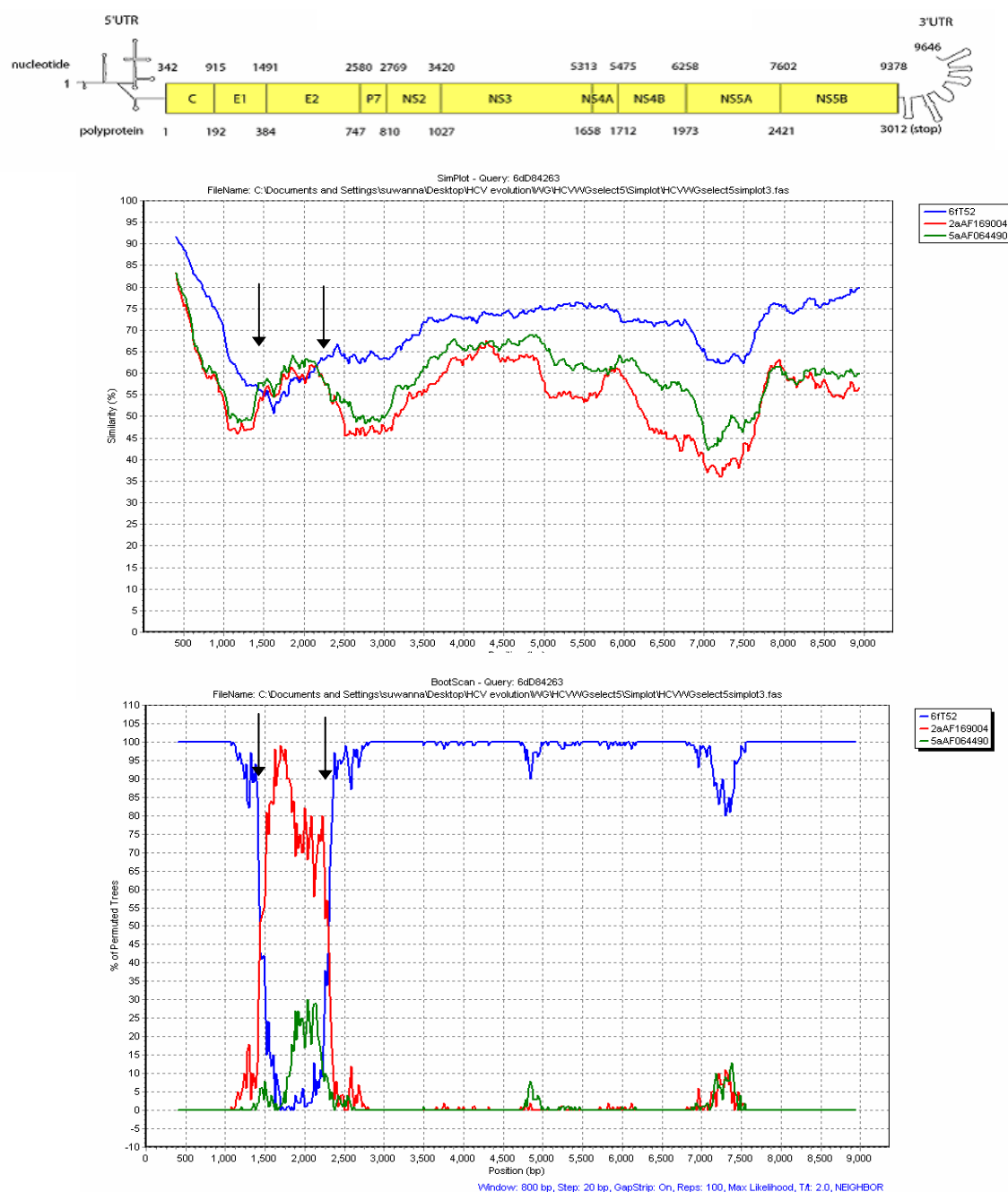


Figure 2. (A) Similarity plots of the intergenotypic recombinant VN235 strain (genotype 6d) and HCV genotype 2a (G2AK3), 6f (TH52) and 5a (SA13) based on the entire genome using a window size of 800 bp, a step size of 20 bp, and the maximum-likelihood parameter. Gaps were ignored, and 1,000 bootstrap replicates were used. (B) Bootscan plots showing the likelihood of clustering of the putative intergenotypic recombinant of VN235 sequence with reference strains, with 1000 bootstrap replicates and neighbor-joining tree analysis. Crossover junctions locate at the nucleotide position 1468-2289. As a reference, the genomic organization of HCV, drawn to scale, is depicted at the top. Arrows indicate the recombination breakpoints. The nucleotide positions are numbered according to the numbering system for the prototype strain H-77 (accession number AF009606, Los Alamos HCV Database)

Table 1 Inter- and intra-genotypic HCV recombinant strains.

Name	Accession number	Recombinant region	Recombinant genotype	Year	Origin	References
N687	AY587845	2k/1b	NS2/NS3	2002	Russia	(18)
PE22	AJ82127	1a/1b	NS5B	2004	Peru	(8)
	AJ582131					
	AJ582134					
SE-03-07-1689	DQ364460	2b/1b	NS3	2006	Philippines	(17)
D3	DQ155560	2i/6p	NS2/NS3	2006	Vietnam	chapter 3
HC-J1	D10749*	1a/1c	E1/E2	2006	Japan	(10)
RF25	AM408911	2/5	NS2/NS3	2007	France	(24)
VN235	D84263**	6f/2a	E1/E2	2008	Vietnam	this thesis

* reported in 1992 as genotype 1a (37)

** reported in 1998 as genotype 6d (56)

We conclude that the assumption that HCV diversity is caused only via genetic drift is no longer tenable and that the role recombination plays in HCV evolution and biology demands serious consideration (Chapter 3). It would be interesting to know if virus chimeras develop more frequently *in vivo*, what type of junction is preferred, and to what extent the appearance of chimeras contributes to pathogenesis, efficacy of therapy and vaccination.

Deletion mutants

During the effort to characterize HCV using phylogenetic analysis of different genome regions, we stumbled upon a sample that apparently contained a double infection of genotype 2i and 6h viruses. Surprisingly, we detected a full-length genome of genotype 2i, and a natural HCV mutant genotype 6h that contained a 674 amino acid in-frame deletion in the E1 and E2 genes (chapter 4). Further analysis showed the presence of deletion mutants in about 10-20% of chronic HCV patients infected with genotype 1, 3 or 6. These HCV mutant RNAs always co-existed with full-length genomes and contained a large in-frame deletion varying from 1254 to 2022 nucleotides in the E1-NS2 region of the genome. The presence of HCV deletion mutants contributes to the already prominent genetic diversity of HCV.

More recently, several HCV deletion mutants have been described in immunosilent chronically infected patients and liver transplant recipients with recurrent

HCV infection (1, 15, 61). All mutant RNAs forms described thus far show a deletion within the E1 and E2 regions (Fig. 3). The deletion genomes contain essential parts for autonomous HCV replication (5'UTR, core, and NS2-NS3-3'UTR), similar to what has been described for the artificially constructed subgenomic replicons (27, 58), and may therefore represent naturally occurring subgenomic replicons. Interestingly, Steinmann et al., (2008) developed a *trans*-packaging system of subgenomic viral RNAs (51) and demonstrated that subgenomic JFH1 replicons lacking the entire core to NS2 coding region are efficiently encapsidated into infectious virus-like particles by using JFH1 helper viruses. Their data indicate that two types of particles are generated: on one hand, replication-competent infectious HCV particles carrying the complete HCV genome and, on the other hand, virus-like particles carrying the subgenomic replicon RNA and thus supporting only single-round infection. The *trans*-complemented HCV particles penetrate target cells through the CD81 receptor and produce particles with a buoyant density comparable to HCV. It is of note that all HCV natural deletion genomes found in our study (Chapter 4) contain complete core coding regions, suggesting that the presence of the core sequence or the production of core protein from the deletion mutant genome is essential for the survival of deletion mutants *in vivo*.

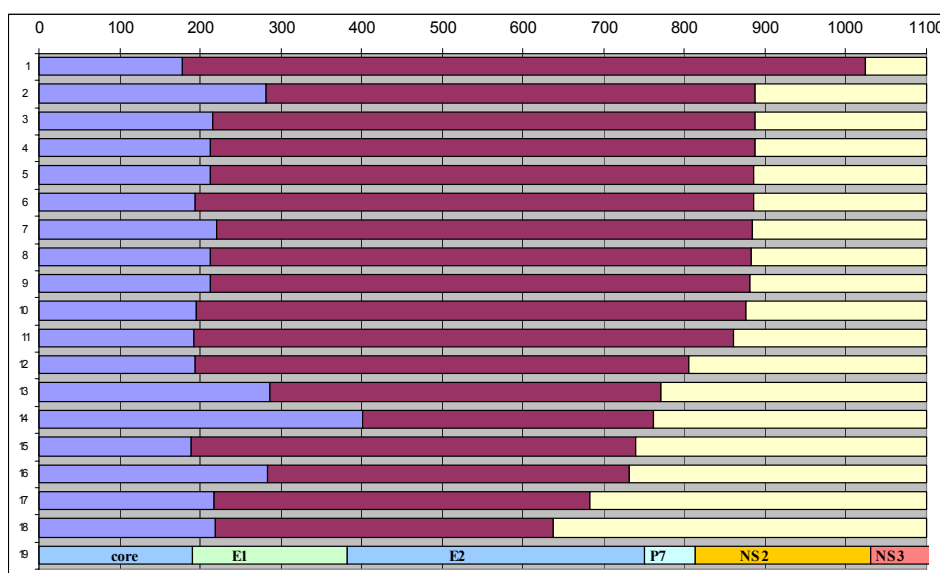


Figure 3 Schematic drawing of HCV deletion mutants described thus far. Numbering at the upper part of the panel refers to the amino acid positions within the HCV genome (acknowledgment to Dr. Giacomo Paonessa, IRBM)

These facts allow us to hypothesize about the generation of deletion mutants during a natural infection with HCV. We suggest that deletion mutant genomes are generated from wild-type HCV within infected cells through errors of the polymerase. In theory, many deletion forms can be generated in the liver (61), however, in-frame deletions which may be favored because of retaining the coding capacity for viral replication, will be selected and circulate. Both the dominant deletion mutant and the wildtype genome supposedly replicate independently. The deletion mutants hijack the structural proteins produced by full-length genomes and produce virus-like particles. At the same time wildtype virus particles are produced as well (Fig. 4a). The ratio of produced wildtype and virus-like particles that is released from infected cells is unknown. It is possible that deletion mutant genomes, due to their reduced genome size, have a replication advantage over full-length genomes, resulting in a higher production of deletion mutant particles than wildtype particles (Fig. 4b). It is also unclear whether coinfection of deletion mutant and wildtype viruses can occur. Deletion mutants can infect target cells and replicate in the absence of full-length HCV RNA (51). Whether core particles are released from these cells is unknown (Fig. 4c). Several studies showed that HCV populations in serum of infected patient are very heterogeneous in size and can be successfully separated by ultracentrifugation (14, 39, 41). Nonenveloped nucleocapsids of HCV were fractionated by CsCl gradient at 1.32 to 1.34 g/ml and showed the ability to bind to immunoglobulin and anti-Core antibodies (30). These observations suggest that deletion mutant genomes may enhance core particle rise into the blood circulation. Infectivity of core particles is unlikely due to the lack of E1/E2 proteins for specific binding protein to the receptor CD81. We cannot exclude, however, that when produced, core particles may spread through other mechanisms, such as cell-to-cell infection, or target other cell types by using a different receptor.

The role of HCV deletion mutants in the life cycle of HCV is unclear. They may play a role in the establishment and maintenance of chronic infection as described for defective interfering RNAs of MVE and JE viruses and their role in persistence *in vitro* (22, 62). Increased levels of HCV Core protein, resulting from increased replication/translation of deletion mutants, may play a role in host immunosuppression, prolonged viremia (23), or disease progression (33). On the other hand, deletion of E1/E2 may have a great impact on the viability of cells either directly through effects on

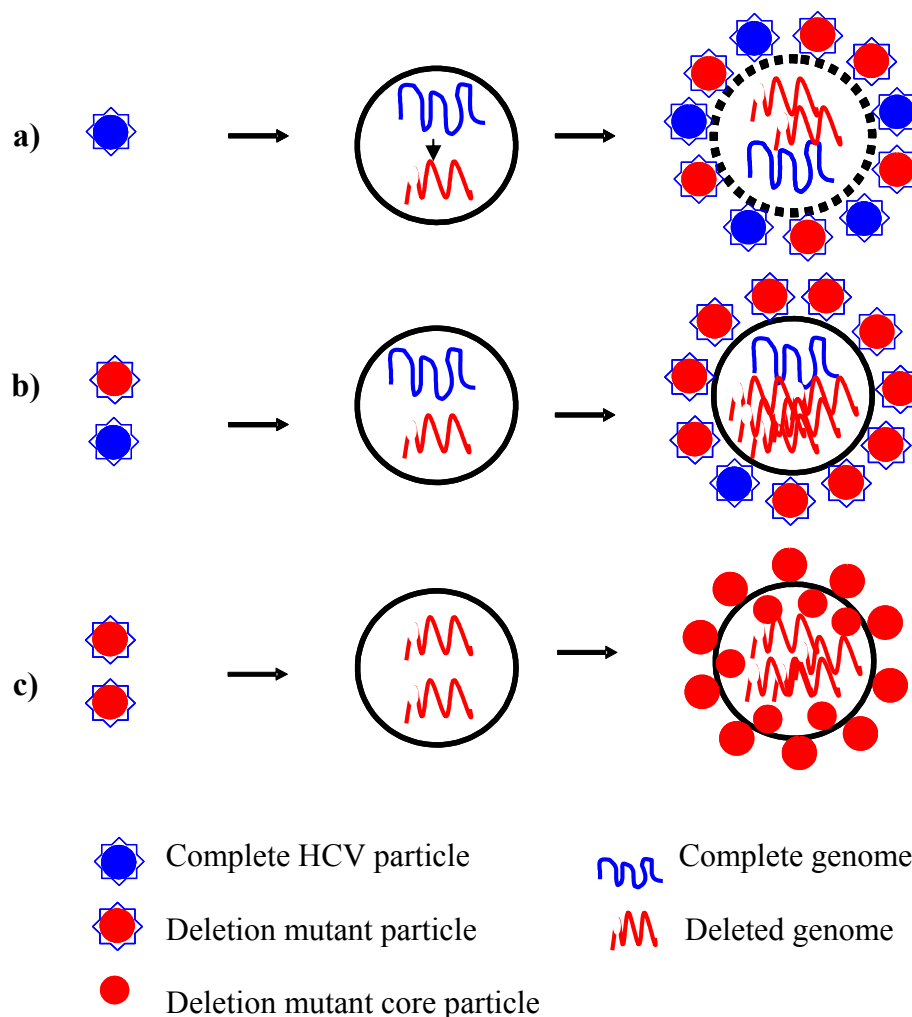


Figure 4. Proposed model of HCV deletion mutant propagation and particle generation.

the ER induced stress response or indirectly by modulating the immune response. Deletion mutants may interfere with the host immune response against HCV, since E1/E2 act as a target molecule for neutralizing antibodies that block the binding of viral particles to host cell receptors (2). Whether and how HCV deletion genomes play a role in HCV infection remains to be delineated. To unravel the role of HCV deletion mutants in HCV biology, methods to discriminate between deletion mutant and wildtype virus need to be developed. This is of particular importance in antiviral therapy studies, where response/non-response is defined by measurements of viral load that are determined based on amplification of the 5'UTR (36), which is the sum of deletion mutant and wildtype virus.

New HCV subtypes 6

Because of the high diversity of HCV genomes, international standardization and coordination of the nomenclature of variants of hepatitis C virus (HCV) is increasingly needed. Recently, a standardized procedure for HCV nomenclature of currently described variants of HCV was proposed and phylogenetic analysis of entire viral genome sequences was set as a golden standard (46). However, genome sequencing of highly diverse strains, such as subtypes of genotype 6, is not always successful due to mismatching of the primers during RT-PCR. We have developed a procedure to amplify large fragments of HCV by RT-PCR from only 100 µl of plasma, using degenerate primers (chapter 3). Based on this new technique, entire genomes of subtype 6a, 6e, 6f, 6i, 6l, 6n and 6o were obtained. Moreover, two new genome sequences both isolated from Vietnamese blood donors were characterized as new subtypes 6t and 6u (chapter 5). Other strains of subtypes 6t and 6u were characterized from Asian immigrants in Canada (34) demonstrating that HCV genotype 6 infection is not restricted to South-East Asia, but spreads throughout the world.

As mentioned before, HCV genotypes show different geographic distributions, with genotypes 1-3 being prevalent worldwide, whereas genotypes 4-6 that show more sequence diversity, are more restricted to African and Asian countries (51). Population genetic models suggest that there are differences in the epidemic history between HCV strains. Growth rates of genotype 1 viruses are higher than those of genotype 4 and 6 viruses, suggesting a recent rapid spread of genotype 1 viruses, in contrast to a long period of localized endemic infections for genotypes 4 and 6. These observations are thought to reflect different modes of transmission. Genotype 1 viruses concurred the world recently through injecting drug use and contaminated blood products, whereas genotype 4 and 6 viruses are thought to spread via undefined social and domestic routes (46). Whereas some genotype 6 viruses, such as subtype 6a, occur throughout South-East Asia, some subtypes are very rare among which 6b, 6c, 6o, 6p, and 6q. Clusters of specific subtypes were found in restricted areas, such as subtype 6c, 6f, 6i and 6j that only circulate in Thailand, subtype 6d, 6h and 6l are found in Vietnam while genotype 6m and 6n are present in Thailand and Myanmar (Fig. 5). Although there may be some sampling bias and information about subtype distribution is lacking from some countries

such as Laos, Cambodia and Myanmar, these results suggest that the spread of HCV genotype 6 is indeed limited to closed populations.

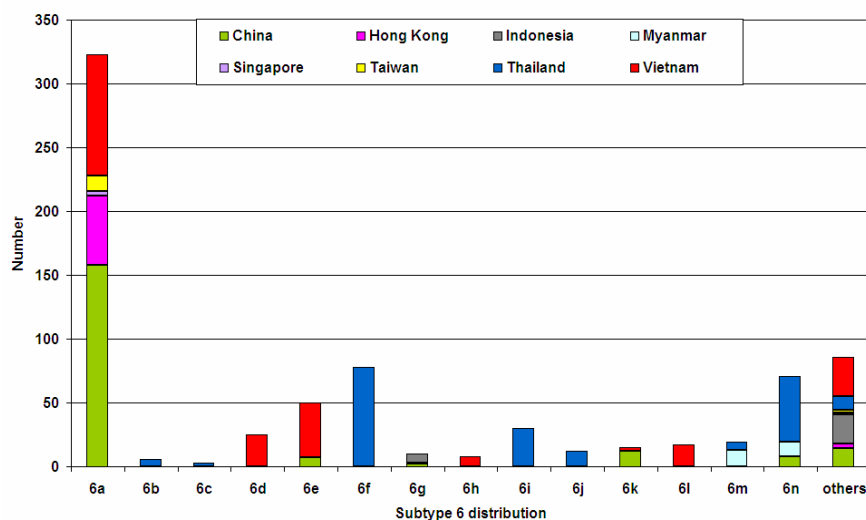


Figure. 5 Deposited HCV genotype 6 sequences from Asian countries.

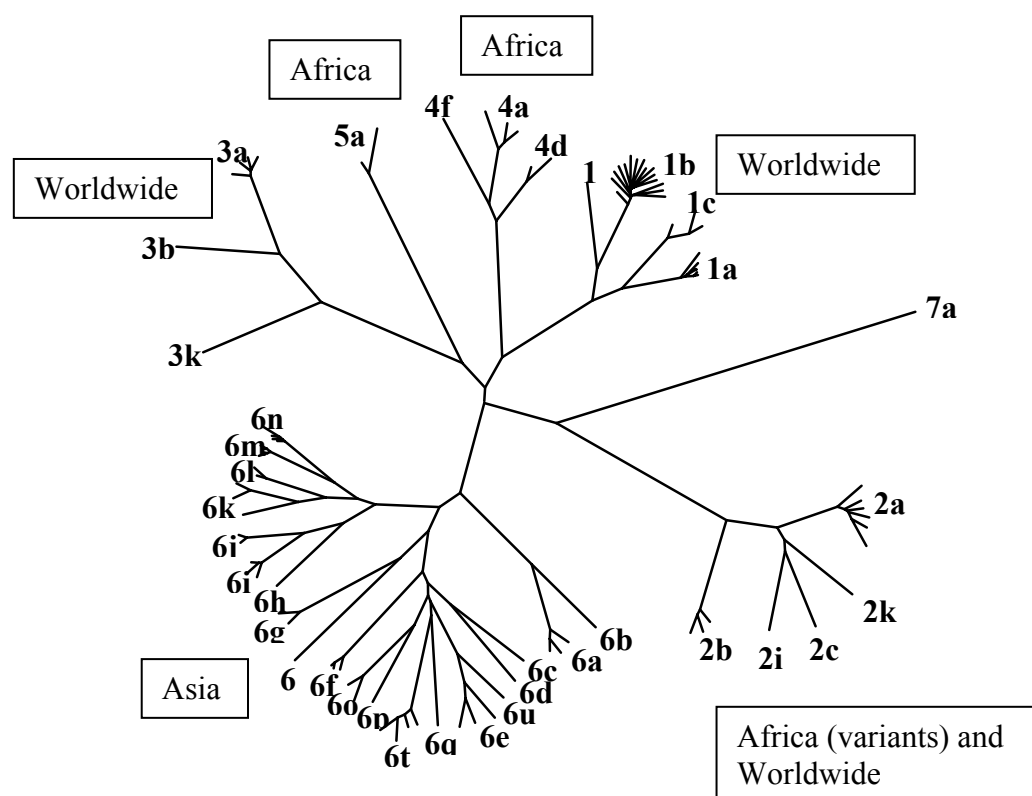
Molecular evolution of HCV

HCV evolution is believed to be mediated by several processes such as genetic drift caused by lack of proof-reading activity in the viral RNA-dependent RNA-polymerase during replication (52), immune pressure (3) and recombination (18). Despite the >30% sequence diversity that is observed among HCV genotypes, they all retain a similar replication cycle in human hosts. Moreover, the ability to establish persistent infection with high circulating viral loads and to produce a slowly progressive and largely asymptomatic infection are similar between different genotypes. Thus, HCV evolution might be affected by neutral evolution with little or no effect of nucleotide changes on organism fitness, resulting in chance fixation of variants in the population (20). ‘Neutral’ sequence drift undoubtedly accounts for much of the genetic diversity that is observed between geographically or epidemiologically separated populations of HCV.

Very little is known about the divergence of the six genotypes of HCV and the origins of this infection in humans. In Africa and Asia, where high divergence strains emerged, human and ape population ranges overlap. The most closely related viruses to HCV, hepatitis GB viruses B (HGV-B), were found to naturally infect tamarins and other new world primates (48). There is, however, no evidence to support cross-species

transmission of HCV from animal to humans, which has hampered dating the origin of the HCV. In chapter 6, we inferred phylogenetic relationships and estimated the times of divergence of the six main clades and also the times of splitting of HCV genotypes, using a relaxed molecular clock model, assuming a nucleotide substitution rate over time. We concluded that the phylogenetic signal of the NS3 region is closest to that of the full-length genome and therefore best to investigate HCV phylogeny for viral identification (both genotyping and subtyping), epidemiology and evolution studies. The estimated rate of nucleotide changing across sites among all genotypes revealed an average high rate of evolutionary change up to 5.71×10^{-4} (full-length genome) to 1.45×10^{-3} (NS3 gene) nucleotide substitutions per site per year, similar to the approximate rate calculated from long term chronic HCV infection in chimpanzees (38). This rate suggested that the age of the entire tree, based on the whole genome sequences, was about 4600 years, more ancient than determined in other studies (50). The most recent common ancestors of genotype 2 and 7, and 3 and 5 would have originated ~4400 and ~4000 years ago, respectively. Surprisingly, we found that genotype 1 and 4 were generated in the same era as genotype 6 in the range of 2600-3000 years ago.

Our data suggest that modern dispersal is responsible for the phylogeographic distribution of HCV (Fig. 6). For example, genotype 1 that is prevalent worldwide, and genotype 4, mainly restricted to Africa, share a common ancestor. Moreover, relatively rare subtypes 1d, 1e, 1g, 1h, 1l and 1m have been isolated from African countries as well, suggesting that genotype 1 may have originated from Africa. A few “founder” strains may have been transmitted rapidly through blood transfusion and sharing of contaminated needles. These strains contain low sequence diversity, resulting from the rapid spreading of a few original strains. Genotype 2 and 6 strains, on the other hand, show high sequence diversity and are basically restricted locally, suggesting a long period of endemic HCV infection in these areas through vertical and horizontal transmission (42, 52). These conclusions are similar to the ones drawn before (46), even though the estimated dates of origin of HCV genotypes were dissimilar.



0.1

Figure 6. The best unrooted ML tree depicting the phylogenetic relationship among 96 HCV strains obtained from HCV database. Prevalence areas of genotypes are indicated.

Detection of HCV specific CD8⁺ T cells

Viral escape typically occurs in the presence of a CTL response that is focused on a single viral epitope. This type of T cell response, however, is unusual during acute HCV infection and accordingly, the loss of a single epitope would probably not be sufficient for the survival of viral escape mutants. It is also important to emphasize that selection of escape variants occurs in the presence of an HCV-specific T cell response that is significantly weaker compared to animals that are able to clear the virus (9, 13). Thus, the different outcomes of infection, clearance versus persistence, may not be explained primarily by the occurrence of viral escape mutations but by initially weak virus-specific T cell responses that make viral escape possible. According to this notion, escape may be the result rather than the cause of viral persistence. Interestingly, escape mutations detected in chronically infected patients did not diversify further during

several years of follow-up (6). Consequently, it was suggested that T cell escape occurs early during infection. This hypothesis has been subsequently confirmed by an analysis of the early T cell–virus interactions in acutely infected chimpanzees (13). Indeed, three animals that developed viral persistence acquired mutations in multiple epitopes that impaired MHC class I binding and/or cytotoxic T lymphocyte (CTL) recognition within the first 16 weeks of infection. As expected from the human studies, these mutations remained fixed for years without further diversification and CD8⁺ T cell responses against variant epitopes were not observed. The failure to generate variant specific CD8⁺ T cell responses can be explained by the lack of sufficient priming or suppression by regulatory T-cells. In this thesis, we have developed an efficient method to detect HCV specific CTL using plasmids encoding NS3 of HCV. These plasmids can be manipulated easily and transfected in BLCL to be used as antigen presenting cells (chapter 7).

Main conclusions

Viral escape mutations are crucial for the development of HCV persistence. The precise mechanisms by which these evolve, however, are unclear. We identified new subtypes 6t and 6u (chapter 5) that probably originated from genetic drift through a long term presence in human hosts. Evidence for genetic shift was obtained by the identification of two recombinant HCV strains, 2i/6p (chapter 3) and 6f/2a (chapter 8). Moreover, the formation of spontaneous deletion mutants, circulating in chronically infected patients, may contribute to the already considerable genetic diversity of HCV (chapter 4). Based on our observations, we would recommend that HCV characterization is best done based on full-length sequences, enabling detection of recombinant strains and deletion mutants. This may be of significant importance as HCV genotype is an important predictor for efficiency of antiviral treatment and progression of hepatitis (62) and deletion mutants may potentially play a role in treatment efficacy as well (15). However, full-length genome sequencing is often not an option for routine HCV genotyping in the clinical setting. Currently, the new VERSANT HCV Genotype 2.0 assay may be the most suitable tool (chapter 2) to discriminate between most subtypes. In addition, mixed infections can be detected using this test. However, one should keep in mind that it cannot detect recombinants. Considering the complexity of HCV genome sequences, the application of phylogenetic analysis tools for characterization and

evolution studies should be considered. Using these tools, we demonstrated that the NS3 region of the genome contains the strongest phylogenetic signal to study the evolution of HCV (chapter 6). Viral escape mutants as described in this thesis may have an important impact on the pathogenesis of HCV. Methods to detect these variants and to analyse the immune response against them (chapter 7) can be used to improve our understanding of the biology of HCV, and the development of antiviral therapies and effective vaccines.

REFERENCES

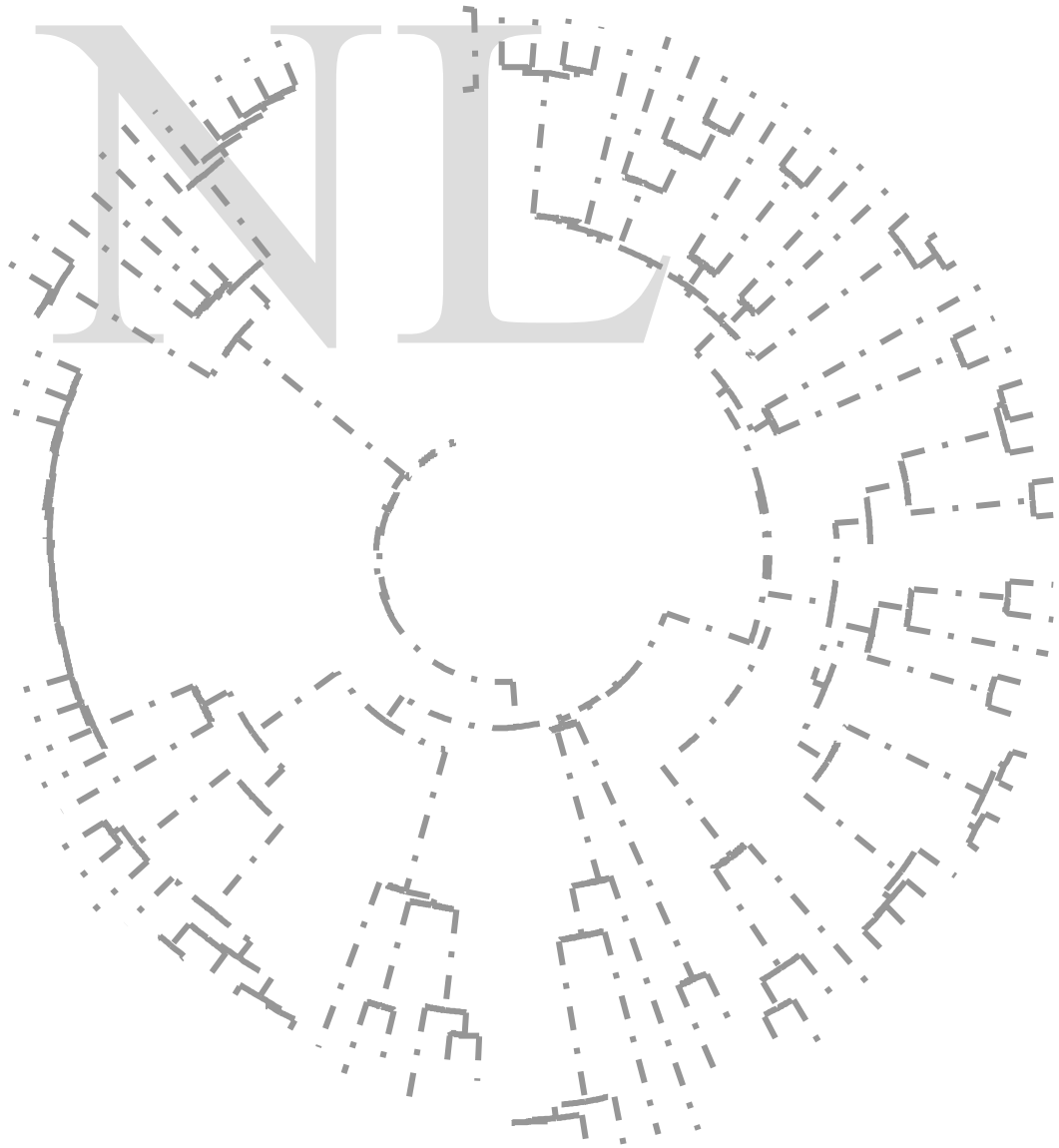
1. **Bernardin, F., S. L. Stramer, B. Rehermann, K. Page-Shafer, S. Cooper, D. R. Bangsberg, J. Hahn, L. Tobler, M. Busch, and E. Delwart.** 2007. High levels of subgenomic HCV plasma RNA in immunosilent infections. *Virology* **365**:446-56.
2. **Beyene, A., A. Basu, K. Meyer, and R. Ray.** 2002. Hepatitis C virus envelope glycoproteins and potential for vaccine development. *Vox Sang* **83 Suppl 1**:27-32.
3. **Booth, J. C., U. Kumar, D. Webster, J. Monjardino, and H. C. Thomas.** 1998. Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. *Hepatology* **27**:223-7.
4. **Bouchardeau, F., J. F. Cantaloube, S. Chevaliez, C. Portal, A. Razer, J. J. Lefrere, J. M. Pawlotsky, P. De Micco, and S. Laperche.** 2007. Improvement of hepatitis C virus (HCV) genotype determination with the new version of the INNO-LiPA HCV assay. *J Clin Microbiol* **45**:1140-5.
5. **Candotti, D., J. Temple, F. Sarkodie, and J. P. Allain.** 2003. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J Virol* **77**:7914-23.
6. **Chang, K. M., B. Rehermann, and F. V. Chisari.** 1997. Immunopathology of hepatitis C. *Springer Semin Immunopathol* **19**:57-68.
7. **Chlabicz, S., R. Flisiak, O. Kowalczyk, A. Wiercinska-Drapalo, B. Pytel-Krolczuk, D. Prokopowicz, and L. Chyczewski.** 2008. High prevalence of genotype 4 among hepatitis C virus-infected intravenous drug users in north-eastern Poland. *J Med Virol* **80**:615-8.
8. **Colina, R., D. Casane, S. Vasquez, L. Garcia-Aguirre, A. Chunga, H. Romero, B. Khan, and J. Cristina.** 2004. Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol* **85**:31-7.
9. **Cooper, S., A. L. Erickson, E. J. Adams, J. Kansopon, A. J. Weiner, D. Y. Chien, M. Houghton, P. Parham, and C. M. Walker.** 1999. Analysis of a successful immune response against hepatitis C virus. *Immunity* **10**:439-49.
10. **Cristina, J., and R. Colina.** 2006. Evidence of structural genomic region recombination in Hepatitis C virus. *Virol J* **3**:53.
11. **Dev, A. T., R. McCaw, V. Sundararajan, S. Bowden, and W. Sievert.** 2002. Southeast Asian patients with chronic hepatitis C: the impact of novel genotypes and race on treatment outcome. *Hepatology* **36**:1259-65.
12. **Doi, H., C. Apichartpiyakul, K. I. Ohba, M. Mizokami, and H. Hotta.** 1996. Hepatitis C virus (HCV) subtype prevalence in Chiang Mai, Thailand, and identification of novel subtypes of HCV major type 6. *J Clin Microbiol* **34**:569-74.
13. **Erickson, A. L., Y. Kimura, S. Igarashi, J. Eichelberger, M. Houghton, J. Sidney, D. McKinney, A. Sette, A. L. Hughes, and C. M. Walker.** 2001. The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. *Immunity* **15**:883-95.
14. **Hino, K., K. Fujii, M. Korenaga, C. Murakami, M. Okazaki, M. Okuda, and K. Okita.** 1997. Correlation between relative number of circulating low-density

- hepatitis C virus particles and disease activity in patients with chronic hepatitis C. *Dig Dis Sci* **42**:2476-81.
15. **Iwai, A., H. Marusawa, Y. Takada, H. Egawa, K. Ikeda, M. Nabeshima, S. Uemoto, and T. Chiba.** 2006. Identification of novel defective HCV clones in liver transplant recipients with recurrent HCV infection. *J Viral Hepat* **13**:523-31.
 16. **Jeannel, D., C. Fretz, Y. Traore, N. Kohdjo, A. Bigot, E. Pe Gamy, G. Jourdan, K. Kourouma, G. Maertens, F. Fumoux, J. J. Fournel, and L. Stuyver.** 1998. Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *J Med Virol* **55**:92-7.
 17. **Kageyama, S., D. M. Agdamag, E. T. Alesna, P. S. Leano, A. M. Heredia, I. P. Abellanosa-Tac-An, L. D. Jereza, T. Tanimoto, J. Yamamura, and H. Ichimura.** 2006. A natural inter-genotypic (2b/1b) recombinant of hepatitis C virus in the Philippines. *J Med Virol* **78**:1423-8.
 18. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius.** 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* **76**:4034-43.
 19. **Kelleher, A. D., C. Long, E. C. Holmes, R. L. Allen, J. Wilson, C. Conlon, C. Workman, S. Shaunak, K. Olson, P. Goulder, C. Brander, G. Ogg, J. S. Sullivan, W. Dyer, I. Jones, A. J. McMichael, S. Rowland-Jones, and R. E. Phillips.** 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J Exp Med* **193**:375-86.
 20. **Kimura, M.** 1983. Rare variant alleles in the light of the neutral theory. *Mol Biol Evol* **1**:84-93.
 21. **Kurbanov, F., Y. Tanaka, D. Avazova, A. Khan, F. Sugauchi, N. Kan, D. Kurbanova-Khudayberganova, A. Khikmatullaeva, E. Musabaev, and M. Mizokami.** 2008. Detection of hepatitis C virus natural recombinant RF1_2k/1b strain among intravenous drug users in Uzbekistan. *Hepatol Res* **38**:457-464.
 22. **Lancaster, M. U., S. I. Hodgetts, J. S. Mackenzie, and N. Urosevic.** 1998. Characterization of defective viral RNA produced during persistent infection of Vero cells with Murray Valley encephalitis virus. *J Virol* **72**:2474-82.
 23. **Large, M. K., D. J. Kittlesen, and Y. S. Hahn.** 1999. Suppression of host immune response by the core protein of hepatitis C virus: possible implications for hepatitis C virus persistence. *J Immunol* **162**:931-8.
 24. **Legrand-Abravanel, F., J. Claudinon, F. Nicot, M. Dubois, S. Chapuy-Regaud, K. Sandres-Saune, C. Pasquier, and J. Izopet.** 2007. New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* **81**:4357-62.
 25. **Leslie, A. J., K. J. Pfafferott, P. Chetty, R. Draenert, M. M. Addo, M. Feeney, Y. Tang, E. C. Holmes, T. Allen, J. G. Prado, M. Altfeld, C. Brander, C. Dixon, D. Ramduth, P. Jeena, S. A. Thomas, A. St John, T. A. Roach, B. Kupfer, G. Luzzi, A. Edwards, G. Taylor, H. Lyall, G. Tudor-Williams, V. Novelli, J. Martinez-Picado, P. Kiepiela, B. D. Walker, and P. J. Goulder.** 2004. HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med* **10**:282-9.
 26. **Lindenbach, B. D., M. J. Evans, A. J. Syder, B. Wolk, T. L. Tellinghuisen, C. C. Liu, T. Maruyama, R. O. Hynes, D. R. Burton, J. A. McKeating, and C. M. Rice.** 2005. Complete replication of hepatitis C virus in cell culture. *Science* **309**:623-6.

27. **Lohmann, V., F. Korner, J. Koch, U. Herian, L. Theilmann, and R. Bartenschlager.** 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**:110-3.
28. **Lu, L., C. Li, Y. Fu, F. Gao, O. G. Pybus, K. Abe, H. Okamoto, C. H. Hagedorn, and D. Murphy.** 2007. Complete genomes of hepatitis C virus (HCV) subtypes 6c, 6l, 6o, 6p and 6q: completion of a full panel of genomes for HCV genotype 6. *J Gen Virol* **88**:1519-25.
29. **Lu, L., C. Li, Y. Fu, L. Thaikruea, S. Thongswat, N. Maneekarn, C. Apichartpiyakul, H. Hotta, H. Okamoto, D. Netski, O. G. Pybus, D. Murphy, C. H. Hagedorn, and K. E. Nelson.** 2007. Complete genomes for hepatitis C virus subtypes 6f, 6i, 6j and 6m: viral genetic diversity among Thai blood donors and infected spouses. *J Gen Virol* **88**:1505-18.
30. **Maillard, P., K. Krawczynski, J. Nitkiewicz, C. Bronnert, M. Sidorkiewicz, P. Gounon, J. Dubuisson, G. Faure, R. Crainic, and A. Budkowska.** 2001. Nonenveloped nucleocapsids of hepatitis C virus in the serum of infected patients. *J Virol* **75**:8240-50.
31. **Micalessi, M. I., C. Gerard, L. Ameye, S. Plasschaert, B. Brochier, and R. Vranckx.** 2008. Distribution of hepatitis C virus genotypes among injecting drug users in contact with treatment centers in Belgium, 2004-2005. *J Med Virol* **80**:640-5.
32. **Moreau, I., S. Hegarty, J. Levis, P. Sheehy, O. Crosbie, E. Kenny-Walsh, and L. J. Fanning.** 2006. Serendipitous identification of natural intergenotypic recombinants of hepatitis C in Ireland. *Virology* **3**:95.
33. **Moriya, K., H. Fujie, Y. Shintani, H. Yotsuyanagi, T. Tsutsumi, K. Ishibashi, Y. Matsuura, S. Kimura, T. Miyamura, and K. Koike.** 1998. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat Med* **4**:1065-7.
34. **Murphy, D. G., B. Willems, M. Deschenes, N. Hilzenrat, R. Mousseau, and S. Sabbah.** 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**:1102-12.
35. **Ndjomou, J., O. G. Pybus, and B. Matz.** 2003. Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *J Gen Virol* **84**:2333-41.
36. **Nolte, F. S.** 2001. Hepatitis C virus genotyping: clinical implications and methods. *Mol Diagn* **6**:265-77.
37. **Okamoto, H., N. Kanai, and S. Mishiro.** 1992. Full-length nucleotide sequence of a Japanese hepatitis C virus isolate (HC-J1) with high homology to USA isolates. *Nucleic Acids Res* **20**:6410.
38. **Okamoto, H., M. Kojima, S. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiro.** 1992. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* **190**:894-9.
39. **Petit, M. A., M. Lievre, S. Peyrol, S. De Sequeira, P. Berthillon, R. W. Ruigrok, and C. Treppe.** 2005. Enveloped particles in the serum of chronic hepatitis C patients. *Virology* **336**:144-53.
40. **Pietschmann, T., A. Kaul, G. Koutsoudakis, A. Shavinskaya, S. Kallis, E. Steinmann, K. Abid, F. Negro, M. Dreux, F. L. Cosset, and R.**

- Bartenschlager.** 2006. Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc Natl Acad Sci U S A* **103**:7408-13.
41. **Prince, A. M., T. Huima-Byron, T. S. Parker, and D. M. Levine.** 1996. Visualization of hepatitis C virions and putative defective interfering particles isolated from low-density lipoproteins. *J Viral Hepat* **3**:11-7.
42. **Pybus, O. G., M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey.** 2001. The epidemic behavior of the hepatitis C virus. *Science* **292**:2323-5.
43. **Richter, S. S.** 2002. Laboratory assays for diagnosis and management of hepatitis C virus infection. *J Clin Microbiol* **40**:4407-12.
44. **Ruggieri, A., C. Argentini, F. Kouruma, P. Chionne, E. D'Ugo, E. Spada, S. Dettori, S. Sabbatani, and M. Rapicetta.** 1996. Heterogeneity of hepatitis C virus genotype 2 variants in West Central Africa (Guinea Conakry). *J Gen Virol* **77 (Pt 9)**:2073-6.
45. **Salemi, M., and A. M. Vandamme.** 2002. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol* **54**:62-70.
46. **Simmonds, P., J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell.** 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**:962-73.
47. **Simmonds, P., J. Mellor, T. Sakuldamrongpanich, C. Nuchaprayoon, S. Tanprasert, E. C. Holmes, and D. B. Smith.** 1996. Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *J Gen Virol* **77 (Pt 12)**:3013-24.
48. **Simons, J. N., T. P. Leary, G. J. Dawson, T. J. Pilot-Matias, A. S. Muerhoff, G. G. Schlauder, S. M. Desai, and I. K. Mushahwar.** 1995. Isolation of novel virus-like sequences associated with human hepatitis. *Nat Med* **1**:564-9.
49. **Smith, D. B., J. Mellor, L. M. Jarvis, F. Davidson, J. Kolberg, M. Urdea, P. L. Yap, and P. Simmonds.** 1995. Variation of the hepatitis C virus 5' non-coding region: implications for secondary structure, virus detection and typing. The International HCV Collaborative Study Group. *J Gen Virol* **76 (Pt 7)**:1749-61.
50. **Smith, D. B., S. Pathirana, F. Davidson, E. Lawlor, J. Power, P. L. Yap, and P. Simmonds.** 1997. The origin of hepatitis C virus genotypes. *J Gen Virol* **78 (Pt 2)**:321-8.
51. **Steinmann, E., C. Brohm, S. Kallis, R. Bartenschlager, and T. Pietschmann.** 2008. Efficient trans-encapsidation of hepatitis C virus RNAs into infectious virus-like particles. *J Virol* **82**:7034-46.
52. **Stumpf, M. P., and O. G. Pybus.** 2002. Genetic diversity and models of viral evolution for the hepatitis C virus. *FEMS Microbiol Lett* **214**:143-52.
53. **Stuyver, L., A. Wyseur, W. van Arnhem, F. Hernandez, and G. Maertens.** 1996. Second-generation line probe assay for hepatitis C virus genotyping. *J Clin Microbiol* **34**:2259-66.
54. **Tallo, T., H. Norder, V. Tefanova, T. Krispin, J. Schmidt, M. Ilmoja, K. Orgulas, K. Pruunsild, L. Priimagi, and L. O. Magnius.** 2007. Genetic

- characterization of hepatitis C virus strains in Estonia: fluctuations in the predominating subtype with time. *J Med Virol* **79**:374-82.
55. **Theamboonlers, A., T. Chinchai, K. Bedi, P. Jantarasamee, M. Sripontong, and Y. Poovorawan.** 2002. Molecular characterization of Hepatitis C virus (HCV) core region in HCV-infected Thai blood donors. *Acta Virol* **46**:169-73.
 56. **Tokita, H., H. Okamoto, H. Iizuka, J. Kishimoto, F. Tsuda, Y. Miyakawa, and M. Mayumi.** 1998. The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9 and comparison with those in the other eight genetic groups. *J Gen Virol* **79 (Pt 8)**:1847-57.
 57. **Tokita, H., H. Okamoto, F. Tsuda, P. Song, S. Nakata, T. Chosa, H. Iizuka, S. Mishiro, Y. Miyakawa, and M. Mayumi.** 1994. Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *Proc Natl Acad Sci U S A* **91**:11022-6.
 58. **Wakita, T., T. Pietschmann, T. Kato, T. Date, M. Miyamoto, Z. Zhao, K. Murthy, A. Habermann, H. G. Krausslich, M. Mizokami, R. Bartenschlager, and T. J. Liang.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med* **11**:791-6.
 59. **Wansbrough-Jones, M. H., E. Frimpong, B. Cant, K. Harris, M. R. Evans, and C. G. Teo.** 1998. Prevalence and genotype of hepatitis C virus infection in pregnant women and blood donors in Ghana. *Trans R Soc Trop Med Hyg* **92**:496-9.
 60. **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.
 61. **Yagi, S., K. Mori, E. Tanaka, A. Matsumoto, F. Sunaga, K. Kiyosawa, and K. Yamaguchi.** 2005. Identification of novel HCV subgenome replicating persistently in chronic active hepatitis C patients. *J Med Virol* **77**:399-413.
 62. **Yoon, S. W., S. Y. Lee, S. Y. Won, S. H. Park, S. Y. Park, and Y. S. Jeong.** 2006. Characterization of homologous defective interfering RNA during persistent infection of Vero cells with Japanese encephalitis virus. *Mol Cells* **21**:112-20.
 63. **Zein, N. N., J. Rakela, E. L. Krawitt, K. R. Reddy, T. Tominaga, and D. H. Persing.** 1996. Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. Collaborative Study Group. *Ann Intern Med* **125**:634-9.



Nederlandse samenvatting

HCV infectie is een belangrijke oorzaak van chronische hepatitis, cirrhose en levercarcinoom. Zo'n 170 miljoen mensen zijn chronisch geïnfecteerd met HCV en dit aantal neemt jaarlijks toe door het ontbreken van effectieve behandelingen en vaccins. Een bekend fenomeen van HCV is de grote genetische variatie van het virus dat problematisch is voor onder andere virus classificatie, diagnose en vaccin ontwikkeling. Daarbij speelt de heterogeniteit van HCV een belangrijke rol in het falen van de immuun respons van de gastheer om het virus op te ruimen en om de gastheer te beschermen voor re- of superinfectie met verschillende HCV varianten. Om de epidemiologie van HCV beter te begrijpen en om effectieve behandelingsmethodes en vaccins te ontwikkelen, is het noodzakelijk om de genetische diversiteit van HCV en de mechanismen die betrokken zijn bij evolutie van HCV verder te karakteriseren.

Bij de start van deze studie was HCV geclassificeerd in 6 genotypen en een groot aantal subtypen. HCV genotypen hebben een verschillende geografische distributie en mate van diversiteit, veroorzaakt door een verschillende epidemiologische geschiedenis en transmissie route. De meeste HCV infecties wereldwijd (~85%) zijn infecties met genotypen 1a/b en 3, voorkomend in Europa, USA, Japan en Rusland. Genotype 2a infecties zijn gerapporteerd in Europa en Japan, terwijl de meer gedivergeerde genotype 2 subtypen in Afrika voorkomen. Genotypen 4 en 5 komen voor in ~4% van de HCV geïnfecteerde populatie, hebben een hoge genetische diversiteit en worden voornamelijk gezien in Afrika, alhoewel verschillende gevallen bekend zijn in Europa. Genotype 6 virussen, prevalent in ongeveer 2% van de HCV geïnfecteerde populatie, zijn genetisch zeer divers en zijn geclassificeerd in 19 subtypen (6a-6s). Epidemiologische data van subtype 6 virussen uit Zuid-Oost Azië zijn gelimiteerd door moeilijkheden rondom collectie van monsters en virale classificatie. Daarom is de natuurlijke historie van genotype 6 virussen in deze regio onduidelijk. In tegenstelling tot genotype 1 virussen, hebben genotype 6 virussen veel meer sequentie heterogeniteit, een beperkte geografische spreiding en lijken relatief zeldzaam. Dit is karakteristiek voor endemische transmissie. Daarbij zijn de huidige HCV detectie systemen ontwikkeld voor het detecteren en classificeren van hoog prevalentie, wereldwijd voorkomende varianten en zijn mogelijk niet in staat om genotype 6 virussen correct te classificeren. Wij hebben de

HCV genotype 6 diversiteit bestudeerd met behulp van serum monsters uit Thailand en Vietnam om inzicht te verkrijgen in HCV epidemiologie en moleculaire evolutie (Hoofdstukken 2-5). De huidige genotyperings- en subtyperingsmethoden zijn verbeterd door gebruik te maken van phylogenetische analyses en een nieuwe “line probe” techniek (Hoofdstuk 2). Evolutionaire mechanismen van HCV genotype 6 zijn bestudeerd (Hoofdstuk 3-4) en nieuwe subtype 6 virussen geïdentificeerd (Hoofdstuk 5). Bioinformatica analyses zijn toegepast om de evolutie van HCV te bestuderen (Hoofdstuk 6) en moleculaire technieken voor het detecteren van CTL activiteit in chronisch HCV geïnfecteerde patiënten (Hoofdstuk 7).

Als eerste stap in de richting van het karakteriseren van HCV diversiteit in Zuid-Oost Azië, hebben we HCV positieve monsters van bloeddonoren uit Bangkok, Thailand en Ho Chi Minh City, Vietnam uit de periode 2000-2002 geanalyseerd. De prevalentie van verschillende HCV genotypen is bepaald met behulp van de nieuwe VERSANT HCV genotype 2.0 assay. Deze test is ontwikkeld om de accuraatheid van HCV genotype/subtype identificatie te verhogen door het toevoegen van Core-specifieke sequenties die verschil kunnen maken tussen genotype 1a, 1b en 6c-1 varianten (Hoofdstuk 2). HCV positieve monsters uit Ho Chi Minh City bevatten genotype 1 (38%), 2 (10.3%), en 6 (51.7%) sequenties, terwijl genotype 1(40%), 3 (45.3%) en 6 (14.7%) werden gevonden in de Thaise populatie (Hoofdstuk 2, 55). Subtypen van deze monsters waren 6a, 6d, 6e, 6f, 6i, 6l, 6n en verschillende ongeclassificeerde of nieuwe subtypen.

Tijdens ons onderzoek met de VERSANT HCV genotype 2.0 assay, werd genotypering van monsters ook gedaan door sequenzen van Core en NS5B fragmenten. Dit leidde tot de identificatie van een HCV recombinant van genotype 2i en 6p (RF2_2i/6p; Hoofdstuk 3). Het recombinatiebreekpunt ligt tussen nucleotiden 3405-3464. Dit correspondeert met de geconserveerde klievingspositie van het NS2-NS3 protease. Het recombinatiebreekpunt van de als eerst gekarakteriseerde intergenotypische recombinant 2k/1b uit St. Petersburg, Rusland, lag ook in de NS2/NS3 regio van het genoom. Recent is de 2k/1b recombinant geïsoleerd uit patiënten in Ierland, Oezbekistan en Estland. Dit suggereert dat deze recombinant succesvol is in verspreiding en kan competieren met wildtype HCV varianten.

Om te bestuderen of meer recombinatie is opgetreden tijdens de diversificatie van genotype 6 virussen, hebben we 96 HCV genomen uit de HCV database, waaronder 35 sequenties van genotype 6 virussen, verzameld. Met behulp van similarity plot en bootscan analyse hebben we een nieuwe intergenotypische HCV recombinant gevonden. Deze was eerder geclassificeerd als subtype 6d in 1998. Isolaat VN235 (accession nummer D84263) is echter een recombinant virus met sequenties van TH52- (genotype 6f, EU246936) en G2AK3-achtige (genotype 2a, AF169004) voorouders. De recombinatiebreekpunten liggen in de E1 en E2 genen (positie 1468-2289), net als in de eerder gevonden intra-genotypische 1a/1c recombinant. Deze recombinatie gebeurtenissen hebben de ontstane recombinant virussen mogelijk een selectief voordeel gegeven in de mogelijkheid om te ontsnappen aan de gastheer immuun respons door het uitwisselen van de E1/E2 regio.

Tijdens het karakteriseren van HCV met phylogenetische analyses van verschillende genoom regio's, zijn we gestoten op een monster dat blijkbaar zowel genotype 2i als 6h virussen bevatte. In dit monster detecteerden wij een geheel genoom van genotype 2i en een natuurlijk mutant genotype 6h dat een 674 aminozuur in-frame deletie bevatte in de E1 en E2 genen (Hoofdstuk 4). Verdere analyse suggereerde dat deletie mutanten voorkomen in 10-20% van de chronisch geïnfecteerde HCV patiënten met genotypen 1, 3, of 6. De HCV mutant RNAs werden altijd gevonden in combinatie met wildtype genomen en bevatten deleties variërend in grootte van 1254-2022 nucleotiden in de E1-NS2 regio van het genoom. HCV deletie mutanten vergroten de toch al prominente genetische diversiteit van HCV.

Om de sterke diversiteit van HCV genomen in kaart te brengen is een standaard procedure voor genotypering/subtypering van HCV varianten gepostuleerd. Hierbij is de phylogenetische analyse van volledige genomen de gouden standaard. Echter, genoom sequentiebepaling van sterk gedivergeerde varianten, zoals subtypen van genotype 6 virussen, is niet altijd succesvol door "mismatching" van oligonucleotiden in de PCR procedure. Wij hebben een methode ontwikkeld om grote fragmenten van HCV te amplificeren met behulp van RT-PCR uit 100 µl plasma, gebruikmakende van gedegeneerde oligo's (Hoofdstuk 3). Met behulp van deze techniek zijn volledige genomen van subtypen 6a, 6^e, 6f, 6i, 6l, 6n en 6o verkregen. Bovendien hebben we twee nieuwe genoomsequenties bemachtigd, subtypen 6t en 6u, uit Vietnamese bloeddonoren

(Hoofdstuk 5). Andere subtype 6t en 6u varianten zijn gevonden in Aziatische immigranten in Canada. Dit suggereert dat genotype 6 infectie niet langer plaatsgebonden is, maar wereldwijd aan het verspreiden is.

De evolutie van HCV lijkt gemedieerd te worden door verschillende processen, waaronder “genetic drift” veroorzaakt door een gemis in “proof-reading”activiteit van het virale polymerase, druk van het immuunsysteem en recombinatie. Er is weinig bekend van de divergentie van de zes HCV genotypen en de oorsprong van infectie in de humane populatie. In hoofdstuk 6 hebben we phylogenetische verwantschappen bepaald, de tijd van divergentie en de tijd van afsplitsen van de zes genotypen geschat met behulp van een relaxed molecular clock model en een constante nucleotide substitutie snelheid in de tijd aannemende. Wij concludeerden dat het phylogenetische signaal van NS3 het dichtst in de buurt komt van het volledige genoom en daardoor de beste regio is voor het bestuderen van HCV phylogenie voor genotypering/subtypering, epidemiologie en evolutie. De geschatte snelheid van nucleotide verandering van alle genotypen liet een hoge gemiddelde snelheid van evolutionaire verandering zien met tot 5.71×10^{-4} (volledige genoom) en 1.45×10^{-3} (NS3 gene) nucleotide veranderingen per positie per jaar. Deze waarden komen overeen met de snelheid bepaald voor HCV in chronische infectie in chimpansees. Bovendien suggereerde deze snelheid dat de leeftijd van de gehele phylogenetische boom ongeveer 4600 jaar is, ouder dan bepaald in andere studies. De meest recente zelfde voorouder van genotype 2 en 7, en 3 en 5 zouden respectievelijk 4400 en 4000 jaar geleden ontstaan zijn. Genotype 1 en 4 zijn ontstaan in dezelfde tijd als genotype 6, ongeveer 2600-3000 jaar geleden.

Onze data suggereren dat moderne spreiding verantwoordelijk is voor de phylogeografische distributie van HCV. Genotype 1 met een wereldwijde prevalentie en genotype 4 dat voorkomt in Afrika delen dezelfde voorouder, bijvoorbeeld. Relatief zeldzame subtypen 1d, 1e, 1g, 1h, 1l en 1m zijn geïsoleerd in Afrika. Dit suggereert dat genotype 1 zich mogelijk vanuit Afrika heeft verspreid. Een aantal “founder” varianten hebben zich mogelijk snel verspreid via bloedtransfusies en het delen van besmette naalden. Deze varianten bevatten weinig sequentie diversiteit, door de snelle spreiding van een paar originele varianten. Genotype 2 en 6 isolaten, daarentegen, hebben hoge sequentie diversiteit en vertonen een lokale distributie in de wereld. Dit suggereert een lange periode van endemische infectie door verticale en horizontale transmissie. Deze

conclusies komen overeen met die van anderen, ook al zijn de geschatte tijden van ontstaan van HCV genotypen verschillend.

Het ontsnappen van virussen aan het immuunsysteem gebeurt voornamelijk in aanwezigheid van een CTL respons die zich richt op een enkel virale epitoom. Dit type T-cel respons is niet gangbaar tijdens acute HCV infectie en daarom zou het verlies van dit epitoom niet voldoende zijn voor het ontsnappen van het virus. Het is van belang te benadrukken dat selectie van “escape” varianten gebeurt in de aanwezigheid van een virus-specifieke T-cel respons die significant zwakker is dan in individuen die het virus opruimen (9,13). Dus de verschillende uitkomsten van infectie, opruiming versus persistentie, kunnen mogelijk primair niet verklaard worden door het ontstaan van escape mutaties, maar door een initieel zwakke virus-specifieke T-cel respons die virale “escape” mogelijk maakt. Naar aanleiding van deze aanname zou escape eerder het gevolg dan de oorzaak zijn van virale persistentie. Escape mutaties in HCV van chronisch geïnfecteerde individuen stapelden zich niet op in de jaren van infectie. Dus werd er geconcludeerd dat T-cel escape plaatsvindt vroeg in infectie. Deze hypothese werd later onderbouwd door een analyse van de vroege T-cel interacties in acut geïnfecteerde chimpansees. Drie dieren, die een virale persistentie ontwikkelden, hadden mutaties in meerdere epitopen, betrokken bij MHC class I binding en/of CTL herkenning, in de eerste 16 weken na infectie. Deze mutaties bleven gefixeerd voor jaren zonder verdere diversificatie en CD8⁺ T-cel responsen tegen gevarieerde epitopen werden niet gezien. Het falen in de productie van variant-specifieke CD8⁺ T-cel responsen kan verklaard worden door een gebrek in “priming” of suppressie door regulerende T-cellen. In dit proefschrift hebben we een efficiënte methode ontwikkeld voor het detecteren van HCV-specifieke CTL door gebruik te maken van vectoren die NS3 van HCV coderen. Deze vectoren zijn makkelijk te manipuleren en te transfecteren in B-lymphocyten die als antigeen presenterende cellen gebruikt kunnen worden (Hoofdstuk 7).

Virale escape mutaties zijn cruciaal voor het ontwikkelen van HCV persistentie, de precieze mechanismen die hierbij betrokken zijn blijven echter onduidelijk. Wij hebben nieuwe subtypes 6t en 6u geïdentificeerd die waarschijnlijk door genetische drift ontstaan zijn veroorzaakt door lange tijd van persistentie in de gastheer (hoofdstuk 5). Bewijs voor genetische shift is verkregen door het identificeren van twee recombinant

hepatitis C virussen, 2i/6p (hoofdstuk 3) and 6f/2a (hoofdstuk 8). Het genereren van deletiemutanten zoals beschreven in hoofdstuk 4 draagt verder bij tot de gigantische genetische variabiliteit van HCV. Gebaseerd op onze waarnemingen bevelen we aan om volledige genoom analyse te doen voor HCV karakterisatie waardoor deletiemutanten en recombinanten kunnen worden gedetecteerd. Deze verdere karakterisatie kan van belang zijn daar is aangetoond dat HCV genetische variatie en met name de genotype 1 en 4 varianten (62) maar ook het aanwezig zijn van deletiemutanten (15) van voorspellende waarde is voor het succesvol verlopen van antivirale therapie en progressie van hepatitis. Het bepalen van de gehele genoomsequentie van HCV is echter geen optie binnen de klinische setting. De nieuwe VERSANT HCV Genotype 2.0 assay is op dit moment de meest geschikte test om de meest voorkomende HCV genotype en subtypes te bepalen (hoofdstuk 2). Dubbel infecties met HCV kunnen met deze test ook worden aangetoond maar recombinant HCV echter niet. Doordat de HCV genomen zo divers zijn is het noodzakelijk de juiste phylogenetische analyse methoden toe te passen. Wij hebben aangetoond dat de NS3 regio uitermate geschikt is om de evolutie van HCV nader te analyseren (hoofdstuk 6). Virale escape mutanten zoals beschreven in deze studie hebben mogelijk een belangrijke rol in de pathogenese van HCV. Methodes om deze varianten en de immuun respons hier tegen te detecteren (hoofdstuk 7) kunnen worden gebruikt om onze kennis betreffende de biologie van HCV en het ontwikkelen van antivirale middelen en vaccins te verbeteren.



ACKNOWLEDGEMENTS

Finally, I am writing the last and most important part of this thesis. This book would not exist without support and encouragement from so many people. During my long travelling from Thailand to Holland and Italy, with laugh and tear, there is not enough space to thank my family, friends, colleagues, etc. I apologize if I don't print all the names here, but your kindness is always printed in my heart.

I would like to express my appreciation to my promoter, Prof. Ab Osterhaus, for his trustful support. Thank you for your patience during the long period of my study, you gave me an opportunity to learn how to be a scientist, and how to survive from it. Endless gratitude is for Dr. Bart Haagmans, my supervisor who did much more than an advisor. Sorry to be your nightmare for many years, however we always shared a very good time together when the last minute mission has been done! Bart, thank you very much to guide me through the scientific thinking process. Many experiments failed, but I become more and more conscious of what the meaning of Science is. You and Astrid are always my source of inspiration and motivation, supporters, friends, family, and more.

This project would not success without the great collaboration of Prof. Yong Poovorawan (Thailand) and Dr. Truong Xuan Lien (Vietnam). The great thankfulness goes to Aj. Yong who is my second father and first planted me in research world; his guidance and encouragement have driven me to the success I have today. Sincere gratitude is extended to Dr. Erwin Sablon and Dr. Kathy De Nys, for a successful collaboration with Innogenetics NV, Gent, Belgium.

My family in Thailand, Pa and Mae, are my moral supports through many years of hard time. Your patience and unlimited love help me to face the World and I have not enough words to express my appreciation for that. My big family, brother and sisters, nieces and nephews, aunts and uncles, thank you for all your love and help throughout these years. I will never reach this part of the thesis without the insistence and support of my god-mother Mrs. Sonda van Donselaar. Almost ten years since the first day I arrived in Holland, Mrs. Sonda is the person who loved, took care, backed up and forced me to go forward to my aims. Many turns almost drove me out of the ring, but she always pulled me back to the right track. Thanks for your love, hugs, support, dinner, cake, coffee and unforgettable times of my life, also with our beloved Chris, Mrs. Bomans,

Yolanda, Elisabeth, Tiny, Aad, Bob, Edward; thank you for your friendship and to be my family in Holland for many years. Of course, I can not deny to thank my husband, who always understands when I have to stay late in the lab, and work a long day during weekend. Luca, grazie amore, senza di te io non avrei raggiunto questo traguardo oggi. Grazie anche alla tua famiglia ed amici che mi hanno accettato in Italia.

Many thanks are expressed to my paranymphs, Saskia and Juthatip who have to do so much work for me to arrange a long distance promotion. Special thanks to Byron (who expected 2 pages of acknowledgements), Penelope (my thesis's prototype), Hans for the often funny discussions after 8 pm, with cookies (for me) and whisky (for them)! Many people that I would like to thank for their friendship and make me feel at home in department of virology, Guus, Georges, Thijs, Annemiek, Anna, Rui, Zahra, Wim, Robert, Simone, Femke, Corine, Jolianne, Willem, Debby, Patrick, Leo, flu groups – Miranda, Sander, Elfje, Theo, Chantal, Tiny, Nella, Monique, Pascal, Emmie, Gerrie, Vincent and Salin. People from diagnostic lab where I started my training; Bert, Jolanda, Edwin, Suzan, Hans, Robin and Georgina, thanks for you support and kindness. My ex-colleagues who encouraged me since the beginning, Sven, Patrick, Ruud, Richard, Jessica, I do hope we will see sometime somewhere.

I would like to thank my Thai friends in Holland who always make me feel close to family, especially during the Thai dinner. Special thanks to P'Noi, my big sister and Don, Nong Meen, P'Chu, M+, Tum, Max, Mor Kiad, P'Tai, P'Mam and many others in Amsterdam. My beloved friends in Thailand, Jee, Pheung, Tim, Preeda and friends from Sor Kor 9, P'Hlin, P'Aum, P'Aoy, P'Nok, Nong Poj; thanks a lot for your friendship and sincerity naka. Grateful to my boss, colleagues and friends in INGM for supporting me through difficult time of preparing this thesis; Sergio, Raffaele, Marco, Mauro, Massimiliano, Rossana, Cristina, Francesco, Milena, Chiara, Jessica I promise I will join you more for lunch.

Thank you all and please join the success of mine, because it comes from you all!

Suwanna

CURRICULUM VITAE

Suwanna Noppornpanth was born on January 5th, 1970 in Bangkok, Thailand. She obtained Master Degree in Industrial Microbiology from Chulalongkorn University, Thailand in 1995. She started her career as a researcher in the National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agent, Bangkok. During 1998-2002, she obtained her PhD program in Medical Microbiology (Virology), Faculty of Medicine, Chulalongkorn University, and received one year training collaboration at Institute of Virology, Erasmus Medical Center, Rotterdam, The Netherlands. After PhD in Thailand, she continued the research work of genetic diversity of Hepatitis C virus in South East Asia under supervision of Professor Dr. A.D.M.E Osterhaus and Dr. Bart L. Haagmans at Erasmus Medical Center, Rotterdam. In 2006, she acquired the postdoctoral position in Novartis, Siena, Italy and nowadays she is a researcher in Fondazione Istituto Nazionale Genetica Molecolare (INGM), Milan, Italy.

PUBLICATIONS

1. **Noppornpanth, S., and Y. Poovorawan.** 1999. Comparison between RT-PCR and rapid agglutination test for diagnosis of human rotavirus infection. *Southeast Asian J Trop Med Public Health* **30**:707-9.
2. **Noppornpanth, S., N. Sathirapongsasuti, V. Chongsrisawat, and Y. Poovorawan.** 2000. Detection of HbsAg and HBV DNA in serum and saliva of HBV carriers. *Southeast Asian J Trop Med Public Health* **31**:419-21.
3. **Chinchai, T., S. Noppornpanth, A. Theamboonlers, V. Chongsrisawat, and Y. Poovorawan.** 2001. Acute posttransfusion hepatitis C: identification of a common hepatitis C virus strain in donor and recipient using polymorphism analysis. *Infection* **29**:40-3.
4. **Noppornpanth, S., T. Chinchai, P. Ratanakorn, and Y. Poovorawan.** 2001. TT virus infection in gibbons. *J Vet Med Sci* **63**:663-6.
5. **Noppornpanth, S., A. Theamboonlers, and Y. Poovorawan.** 2001. Predominant human rotavirus genotype G1P[8] infection in infants and children in Bangkok, Thailand. *Asian Pac J Allergy Immunol* **19**:49-53.
6. **Chinchai, T., J. Labout, S. Noppornpanth, A. Theamboonlers, B. L. Haagmans, A. D. Osterhaus, and Y. Poovorawan.** 2003. Comparative study of different methods to genotype hepatitis C virus type 6 variants. *J Virol Methods* **109**:195-201.
7. **Noppornpanth, S., B. L. Haagmans, P. Bhattarakosol, P. Ratanakorn, H. G. Niesters, A. D. Osterhaus, and Y. Poovorawan.** 2003. Molecular epidemiology of gibbon hepatitis B virus transmission. *J Gen Virol* **84**:147-55.
8. **Keawcharoen, J., K. Oraveerakul, T. Kuiken, R. A. Fouchier, A. Amonsin, S. Payungporn, S. Noppornpanth, S. Wattanodorn, A. Theamboonlers, R. Tantilertcharoen, R. Pattanarangsarn, N. Arya, P. Ratanakorn, D. M. Osterhaus, and Y. Poovorawan.** 2004. Avian influenza H5N1 in tigers and leopards. *Emerg Infect Dis* **10**:2189-91.
9. **Keawcharoen, J., A. Amonsin, K. Oraveerakul, S. Wattanodorn, T. Papravit, S. Karnda, K. Lekakul, R. Pattanarangsarn, S. Noppornpanth, R. A. Fouchier, A. D. Osterhaus, S. Payungporn, A. Theamboonlers, and Y.**

- Poovorawan.** 2005. Characterization of the hemagglutinin and neuraminidase genes of recent influenza virus isolates from different avian species in Thailand. *Acta Virol* **49**:277-80.
10. **Pas, S. D., S. Noppornpanth, A. A. van der Eijk, R. A. de Man, and H. G. Niesters.** 2005. Quantification of the newly detected lamivudine resistant YSDD variants of Hepatitis B virus using molecular beacons. *J Clin Virol* **32**:166-72.
 11. **Chinchai, T., S. Noppornpanth, K. Bedi, A. Theamboonlers, and Y. Poovorawan.** 2006. 222 base pairs in NS5B region and the determination of hepatitis C virus genotype 6. *Intervirology* **49**:224-9.
 12. **Noppornpanth, S., T. X. Lien, Y. Poovorawan, S. L. Smits, A. D. Osterhaus, and B. L. Haagmans.** 2006. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J Virol* **80**:7569-77.
 13. **Noppornpanth, S., E. Sablon, K. De Nys, X. L. Truong, J. Brouwer, M. Van Brussel, S. L. Smits, Y. Poovorawan, A. D. Osterhaus, and B. L. Haagmans.** 2006. Genotyping hepatitis C viruses from Southeast Asia by a novel line probe assay that simultaneously detects core and 5' untranslated regions. *J Clin Microbiol* **44**:3969-74.
 14. **Noppornpanth, S., S. L. Smits, T. X. Lien, Y. Poovorawan, A. D. Osterhaus, and B. L. Haagmans.** 2007. Characterization of hepatitis C virus deletion mutants circulating in chronically infected patients. *J Virol* **81**:12496-503.
 15. **Noppornpanth, S., Y. Poovorawan, T. X. Lien, S. L. Smits, A. D. Osterhaus, and B. L. Haagmans.** 2008. Complete genome analysis of hepatitis C virus subtypes 6t and 6u. *J Gen Virol* **89**:1276-81.
 16. **Noppornpanth, S., B. Martina, Y. Poovorawan, T.X. Lien, S.L. Smits, A.D. Osterhaus, and B. L. Haagmans.** Molecular evolution of Hepatitis C Viruses: dating the origin of viruses using phylogenetic and relaxed molecular clock analysis. Manuscript preparation.
 17. **Noppornpanth, S., T. Chinchai, S. Bruijns Y. Poovorawan, A. D. Osterhaus, and B. L. Haagmans.** Detection of hepatitis C virus specific T cells using Epstein Barr virus based plasmids. Manuscript in preparation