African Journal of Biotechnology Vol. 9(36), pp. 5792-5799, 6 September, 2010
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2010 Academic Journals

*Review*

# Phylogenetics of HCV: Recent advances

**Nazish Bostan, Maleeha Maria Mustafa, Waseem Safdar, Qamar Javed and Tariq Mahmood**

Department of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan.

Accepted 25 March, 2010

**Hepatitis C virus (HCV), a virus present in human population from indefinite time period, has affected millions of people globally, by causing liver infection which in majority of cases leads to chronicity, cirrhosis, end stage liver disease and hepatocellular carcinoma (HCC). The disease burden is expected to increase in the developing and under developed world in future. The distribution of HCV genotypes is changing, as are the modes of transmission. Evolution of HCV is a highly dynamic process as it exploits all known mechanisms of genetic variation including recombination and mutation, to ensure its survival. It occurs both through multiple processes of adaptive selection that drive sequence change and through drift, in which phenotypically neutral sequence changes accumulate over time without altering the phenotype or behaviour of the virus. However, despite its potential to change rapidly, the longer-term evolution of HCV appears to be remarkably conservative. Phylogenetic and statistical models of viral evolution are useful in reconstructing mutational pathways of drug resistance. The two major divisions of viral heterogeneity include genotypes and quasispecies. The rate of nucleotide changes varies significantly among the different regions of the viral genome. The present HCV classification is incomplete, as new genotypes and variants are being identified till yet. Diversification of HCV occurred over time but with different rates. Host immune pressure is thought to be a main factor driving diversification in HCV quasispecies. Core and hypervariable regions are more diverse while 5' un-translated region (UTR) and 3' UTR are highly conserved across the genotypes.**

**Key words:** HCV, phylogeny, 5' UTR, viral evolution, recombination, quasispecies.

## INTRODUCTION

As early as 752, Pope Zacharias wrote in a letter to St. Boniface, Bishop of Mainz (Germany), about, “Jaundice of a contagious nature” where those affected would have to be segregated. Hippocrates described a potentially dangerous disease widely witnessed in young people and accompanied by jaundice (Kuntz and Kuntz, 2006). Hepatitis C virus is an important human pathogen, not only because of its high prevalence and world wide burden but also because of the potentially serious complications of persistent HCV infection. These complications include cirrhosis, hepatocellular carcinoma and end-stage liver disease. The incident rates of all these complications are expected to rise in the near future. Recognition of the importance of HCV infection has led to the consensus conferences in both the United States and Europe to establish standards of care for HCV-infected patients and to set research priorities. Despite substantial progress, a precise understanding of the mechanisms of HCV replication and persistence and of the pathogenesis of the liver disease caused by HCV, remains elusive. The limited progress in these important areas stems in large part from the lack of efficient cell culture system, as well as the narrow host range of the virus, with few animal models available for study (Dienstag and Isselbacher, 2005).

*Corresponding author. E-mail: taariq.mahmood@yahoo.com.

**Abbreviations**: **HCV**, Hepatitis **C** virus; **HCC**, hepatocellular carcinoma; **UTR**, 5' un-translated region; **IRES**, internal ribosome entry site; **HVR**, hypervariable region; **PCR**, polymerase chain reaction; **ELISA**, enzyme-linked immunosorbent assay; **ORF**, open reading frame; **dS**, synonymous nucleotide substitutions per synonymous site; **dN**, nonsynonymous nucleotide substitutions per nonsynonymous site; **5'NCR**, 5' non-coding region; **MSM**, male sex with male; **CTL**, cytotoxic T-lymphocyte.

## EVOLUTION OF HCV

HCV evolution is a highly dynamic process. HCV exploits all known mechanisms of genetic variation, such as recombination and mutation, to ensure its survival. It

circulates *in vivo* as a complex population of different but closely related variants (Cristina, 2005). Modern HCV antiviral therapies have been successful at keeping virus suppressed for prolonged periods of time, but therapy failures attributable to the emergence of drug resistant mutations continue to be a distressing reminder that no therapy can fully eradicate these viruses from their host organisms. Phylogenetic and statistical models of viral evolution are used to reconstruct mutational pathways of drug resistance. Mutational pathways facilitate the identification of fast versus slow evolutionary pathways to drug resistance (Buendia et al., 2009).

The genomic variability of HCV is due to the high error rate of the viral RNA-dependent RNA polymerase, which has been calculated to be in the range of about $10^{-4}$. This number is in line with the mutation rates of $1.44 \times 10^{-3}$ and $1.92 \times 10^{-3}$ base substitutions per site per year that were found in a chronically infected human and chimpanzee. This phenomenon is favoured by a high viral turnover rate. The rate of production of virion is on the order of $10^{10}$ to $10^{12}$ virions per day (Kronke et al., 2004). A substantial proportion of newly synthesized viral genomes are defective because of highly deleterious genetic lesions, whereas other genetic alterations confer advantages that lead to the continuous selection of these fittest variants under new environmental conditions.

## DIVISONS OF VIRAL HETEROGENETY

The first division used to describe the genetic heterogeneity of HCV is the viral genotype, which refers to genetically distinct groups of HCV isolates that has arisen during viral evolution. Sequencing of nucleotides has shown a 34% variation between the genotypes. The most conserved 5' UTR (untranslated region) has a maximum nucleotide sequence divergence of 9% between the genotypes, whereas the highly variable regions that encode the putative envelop proteins (E1 and E2) exhibit a nucleotide sequence divergence of 35 - 44% between genotypes. The 5' UTR of HCV-RNA forms a secondary structure that guides the ribosomes for translation. This secondary structure is termed the internal ribosome entry site (IRES). The sequences cluster into six major genotypes (Designated by Arabic Numbers), with sequence similarity of 60-70% and more than 50 subtypes (Designated by a lower case letter) within these major genotypes with sequence similarity of 77-80% in this scheme (Pawlotsky, 2003, 2009). Global geographical differences exist in the distribution of HCV genotypes, as well as in the mode of acquisition and response to antiviral therapy. The distribution of HCV genotypes is changing, as are the modes of transmission, however and since 1995, genotypes 1a and 3a, which are common in injection drug uses, have increased in frequency compared with genotype 1b which is associated most commonly with transmission by blood transfusion (Wesley and Alter, 2000). The reasons for the differences in treatment susceptibility are not fully understood but may be attributable in part to genotype specific differences in the interaction between certain HCV protein, such as NS5a and E2 and intracellular pathways that mediate the effects of interferon. A consistent association has been found between HCV genotype 3 infections and liver steatosis, suggesting that HCV genotype 3 may interrupt intracellular metabolism of lipids (Lonardo et al., 2004).

Second component of genetic heterogeneity is Quasispecies. These are closely related, yet heterogeneous sequences of HCV within a single infected person and result from the mutations that occur during viral replication. The rate of nucleotide changes varies significantly among the different regions of the viral genome. The highest rate of mutation has been found in E1 and E2 region, particularly in the hypervariable region (HVR) at the amino-terminal end of E2 (E2HVR1). The present classification of HCV is incomplete, as information from a number of isolates is still awaited (Farci et al., 2000). Phylogenetic analysis of sequence information obtained from the 5'NCR correlates fairly well with that obtained from the core and the NS3, NS4 and NS5 regions of the viral genome in the determination of major genotypes (Germer and Zein, 2001). However, insufficient sequence variation in the 5'NCR limits its usefulness for determining differences among various subtypes (Bukh et al., 1993, 1995). The identification and characterization of HCV types and subtypes have major implications for HCV diagnostics as well as for the development of vaccine.

Variants of HCV have been classified by nucleotide sequence comparisons in different regions of the genome. Many investigators have defined the ranges of sequence similarity values or evolutionary distances corresponding to divisions of HCV into types, subtypes and isolates. Using these criteria, novel variants of HCV from Vietnam, Thailand and Indonesia have been classified as types 7, 8, 9, 10 and 11, many of which can be further subdivided into between two to four subtypes (Simmonds, 2004). Most methods for direct HCV genotyping include amplification of different genome regions, such as the 5'UTR, core, E1 or NS4, by PCR with type-specific primers or by restriction fragment length polymorphism analysis of PCR products (Ohno et al., 1997; Okamoto et al., 1993; Stuyver et al., 1993, 1994). Indirect HCV genotyping may be achieved by demonstration of type-specific antibodies by ELISA (Dixit et al., 1995; Simmonds et al., 1993). Thus, Present methods of HCV genotype identification do not take recombination into account. A typing method based on the sequence of the core region could reliably identify subtypes as well as major genotypes since the sequence divergence was greater than the divergence of the 5'NCR sequence. Though the preferred regions for genotyping are E1, NS4, and NS5, which are more variable than the core, it is not always possible to amplify them. The main reason is a lack of conservation in the primer binding sites (Kavita et al., 2003). The viral RNA

genome is translated by the host translation apparatus as a single polyprotein product, which is then subjected to proteolytic processing to produce the viral proteins. The length of the open reading frame (ORF) of each genotype is characteristically different. For example, the open reading frame in type 1 isolates is approximately 9,400 ribo-nucleotides in length, while that of type 2 isolates is typically 9,099 nucleotides and that of type 3 isolates is typically 9,063 nucleotides. There is published suggestion that the evolutionary (phylogenetic) relatedness between different genotypes should be reconsidered and the number of recognized genotypes into which HCV isolates are classified should be reassessed. Some reports suggest that subsets of HCV genotypes are more closely related to each other than to other more distantly related genotypes, which should be reflected in a modified HCV nomenclature. It is suggested that the 11 genotypes can be regrouped into six HCV clades. The grouping of clades reflects phylogenetic relationships between the genotypes, where genotypes 1, 2, 4 and 5 all represent distinct clades, but where genotypes 3 and 10 are placed into a single lade 3 and genotypes 6, 7, 8, 9 and 11 are placed into a single clade 6 (Zein, 2000).

## TIME ESTIMATION OF GENOTYPE EVOLUTION

Tanaka (2006), studied the coalescent analysis indicated that a transition from constant size to rapid exponential growth (spread time) occurred in Japan in the 1920s (HCV-1b), but not until the 1940s for the same genotype in Spain and other European countries. The spread time of HCV-1a in the United States was estimated to be in the 1960s; HCV-3a in the FSU, HCV-5a in South Africa and HCV-6a in Hong Kong in the 1960s, mid-1950s and late 1970s, respectively. Three different linear progression curves were determined by analysis of the relationship between HCV seroprevalence and HCC mortality in different geographic regions; a steep ascent indicated the greatest progression to HCC in Japan, a near horizontal line indicated the least progression in the United States and the FSU and an intermediate slope was observed in Europe, suggesting that the initial spread time of HCV is associated with the progression dynamics of HCC in each area, irrespective of genotype.

Mizokami and Tanaka (2005), found that the molecular clock has been a very powerful tool in looking back at the epidemic spread of HCV infection in the United States (US) and Japan, as well as in Egypt. This analysis estimates that the growth of the US HCV genotype 1a (HCV-1a)-infected population occurred around 1960, at least 30 years later than the widespread introduction of HCV-1b into the Japanese population. In Japan, the estimated effective number of HCV infections indicated a rapid exponential growth in the 1920s among patients with schistosomiasis, which coincides with injection treatment for schistosomiasis, since 1921 in previously schistosomiasis-endemic areas. In Egypt, the spread of HCV-4a would have increased exponentially during the 1940s through 1980, which was also consistent with the duration of intravenous antimony campaigns for the treatment of shistosomiasis in that country. The implications are that Japan has set the model for HCV-related HCC and that the high HCC incidence in Japan might be replicated by the rest of the world as their HCV-infected population ages and the duration of HCV infection approaches that currently observed in Japan. Evolution of the quasispecies is hypothesized to be due to ongoing selection of viruses that are most fit for a particular host. Selective pressure can be related to several factors but host immune pressure is thought to be a main factor driving diversification. Failing to detect a correlation between HCV evolution and the strength of the host immune response to HVR-1 epitopes, (Allain et al., 2008) have suggested that variation can result from genetic drift occurring independently of immune pressure. Comparison frequencies of synonymous nucleotide substitutions per synonymous site (dS) and nonsynonymous nucleotide substitutions per nonsynonymous site (dN) can be used to evaluate the process of natural selection. In the absence of selection, dS exceed dN in most protein-coding genes. A pattern of dN > dS suggests positive selection (Henry et al., 2004; Honardoost et al., 2008).

The molecular epidemiology of HCV genotype 2 in its region of endemic origin, west and central Africa was analysed including 56 new and highly diverse HCV isolates sampled from infected individuals in Guinea-Bissau. Eastwards spatial spread from the West African coast to Cameroon that took place over several centuries. Molecular clock analysis dates the common ancestor of HCV in Guinea-Bissau to 1470 (1414-1582). The phylogenetic position of isolates from Madagascar and Martinique suggests a role for the historical slave trade in the global dissemination of HCV and of the epidemic subtypes 2a and 2c. Coalescent-based estimates of epidemic growth indicate a rapid 20th-century spread of HCV genotype 2 in Cameroon that is absent in Guinea-Bissau. This contrast is to be discussed in the context of possible parenteral HCV exposure during public-health campaigns undertaken during the colonial era (Markov et al., 2009).

Despite the sequence diversity of HCV, all genotypes share an identical complement of collinear genes of similar or identical size. However, contrasting with this general observation is the marked variation in their capability to express a further protein that is generated by translational frameshift at codon 11 of the core gene (Walewski et al., 2001; Xu et al., 2001; Varaklioti et al., 2002) both the frameshift site and potential size of this novel coding sequence are very poorly conserved between and within genotypes. This contrast with the evolutionarily conserved nature of so many other aspects of HCV replication supports the idea that this ‘gene’ is more likely to be a computational artefact that has arisen from RNA

structure-imposed constraints on third-codon position variability in the core gene (Tuplin et al., 2004).

## EVOLUTION OF 5'UTR REGION OF HCV GENOME

During the last decade, there has been a dramatic increase in intravenous drug use in young adults in Estonia with an increased incidence of hepatitis C and other viral infections, as a consequence. Since genetic data are limited regarding HCV strains in Estonia, the aim of the study was to characterize HCV strains in different risk groups to determine their relatedness to strains from other geographical regions. Three hundred fifty-three anti-HCV positive sera collected during 1994 - 2004 from hospitalized patients, blood donors and health care workers were used as source of HCV RNA. Positive sera for HCV RNA by PCR directed to the 5'-UTR region. HCV subtype was determined by analyses of the NS5B and/or the 5'UTR-core regions. 1b was the most common subtype followed by 3a, 2c, 1a and 2a. The 1b and 3a strains were similar to strains from other regions of the former USSR. Within genotype 1b there were several HCV lineages. However, for 3a there seemed to be two separate introductions into Estonia. There was a relative shift from subtype 1b to 3a in 1999 - 2000 with a further replacement of 3a with 1b in intravenous drug users in 2001 and onwards ($P < 0.05$). However, both subtypes were found to co-circulate in the community independent of risk factors. One patient was infected with the 2 k/1b recombinant presumed to originate from St. Petersburg being the first isolate of this recombinant recovered outside Russia (Tatjana et al., 2007).

Previous and recent studies have suggested a diversification of type 1 HCV in the South American region. The degree of genetic variation among HCV strains circulating in Bolivia and Colombia is currently unknown. In order to get insight into these matters, a phylogenetic analysis was performed of HCV 5' non-coding region (5'NCR) sequences from strains isolated in Bolivia, Colombia and Uruguay, as well as available comparable sequences of HCV strains isolated in South America (Moratorio et al., 2007). Phylogenetic tree analysis of HCV strains isolated in the South American region revealed the presence of a distinct genetic lineage inside genotype 1. Signature pattern analysis revealed that the presence of this lineage is consistent with the presence of a sequence signature in the 5'NCR of HCV strains isolated in South America. Comparisons of these results with the ones found for Europe or North America revealed that this sequence signature is characteristic of the South American region. Phylogentic analysis revealed the presence of a sequence signature in the 5'NCR of type 1 HCV strains isolated in South America. This signature is frequent enough in type 1 HCV popula-tions circulating South America to be detected in a phylogenetic tree analysis as a distinct type 1 sub-population. The co-existence of distinct type 1 HCV sub-populations is consistent with quasispecies dynamics and suggests that multiple coexisting subpopulations may allow the virus to adapt to its human host populations (Roque-Afonso, 2005; Moratorio et al., 2007). A significant proportion of HCV-infected subjects harbour in their peripheral blood mononuclear cells highly divergent variants which were likely acquired through super-infections (Roque-Afonso, 2005).

Forty five recent isolates of Hepatitis C virus (HCV) originating from four different geographic regions of the world, were typed using phylogenetic analysis of a 192 nucleotides (nts) long sequence from the 5'non-coding region (5'-NCR) of the virus genome and compared them with 55 HCV isolates/strains of known type. During the study it was observed that phylogenetic studies can assign an HCV isolate to the correct type in 100% and to the correct subtype in 98%. A comparison of this method with other methods using commercial kits revealed that it is appropriate for clinical use and is cost effective (Cristina et al., 2002). Investigating the evolution of the HCV genome in the small number of patients that experience viral breakthrough might shed light on the problem of resistance to interferon therapy. Within the HCV genome, sequence diversity of the viral nonstructural 5A protein-coding region (NS5A) has been linked to interferon responsiveness. Comparing baseline sequences showed substitutions were focused in the V3 and flanking regions in Break Through patients but not in Sustained Virological Response patients. The high number of substitutions in NS5A in both Break Through and Sustained Virological Response groups suggests that selective pressure is associated with viral response to therapy. Amino acid substitutions within the NS5A coding region may reflect a host response that drives selective pressure for viral adaptation (Yuan et al., 2009).

## EVOLUTION OF CORE REGION OF HCV GENOME

Phylogenetic analysis of core region of HCV genome revealed a large international network of HCV transmission among HIV-positive MSM (Male Sex with Male). The rapid spread of HCV among neighboring countries is supported by the large proportion (74%) of European MSM infected with an HCV strain co-circulating in multiple European countries, the low evolutionary distances among HCV isolates from different countries and the trend toward increased country mixing with increasing cluster size. Temporally, this epidemic coincides with the introduction of highly active antiretroviral therapy and associated increases in sexual risk behaviours (Laar et al., 2009).

The most extensively documented area is an 81-bp region in the N-terminus of the envelop 2 region called the hypervariable region 1 (HVR-1), a major target for neutralizing antibodies. Analysis of HVR-1 allows establishment of linkage between different sequences. Perinatal

transmission can be verified by demonstrating close linkage between maternal and infant HCV isolates. Weiner et al., (1993) reported a case of an HCV-infected newborn whose unique isolate was different from those of the mother and raised the issue of whether the transmitted virus was an escape mutant or whether selection occurred at the time of transmission. Subsequent studies in children have demonstrated diverse patterns of transmission including transmission of multiple clones, transmission of dominant or subdominant clones, or a mixture of these (Henry et al., 2004).

Diversification of HCV occurred over time but with different rates. Differential HCV phylogenetic evolution of common source-infected individuals suggests that individual host selective pressures are at play in determining quasispecies transmission and evolution. The effect of immune pressure, or adaptive selection, not only increases viral diversification but may also result in escape from humoral or cellular immunity. A number of studies have documented that the hypervariable region (HRV-1) in the E2 region of the virus is a dominant neutralization epitope. The carboxyterminal end of HRV-1 also contains epitopes for both T-helper and cytotoxic responses. The presence of immune pressures could play a role in determining which viral clones are passed from mother to child and could also influence the evolution of the viruses in both the mother and child over time. Additional factors involved might be the mode of transmission and potentially different quasispecies populations, that is, blood, placenta vs. blood, vaginal fluid and the amount of inoculums, etc. The high rate of changes in the 1st year of life, a time when infant immunity is generally diminished, reflects possible viral escape as a result of declining titers of passively acquired maternal HCV antibody rather than newly acquired HCV-specific immune responses in the infant. After the 1st year of life, changes in infant quasispecies populations are less dramatic and suggest a more slowly evolving process (Henry et al., 2004).

## EVOLUTION OF X-TAIL 3'UTR OF HCV GENOME

Drexler et al. (2009), determined by de novo sequencing that the 3′-X-tail element, characterized significantly later than the rest of the genome, is highly conserved across genotypes. A prototype qualitative and quantitative test was developed and evaluated multicentrically on a large and complete panel of clinical plasma samples, to prove its clinical utility as a molecular diagnostic target, covering HCV genotypes 1-6, from four continents namable Germany, UK, Brazil, South Africa and Singapore. In recent past HCV X-tail region has been targeted for molecular detection and quantification, in order to circumvent the limitations that the 5′-NCR presents in diagnostics. But its utility was not sure because of several reasons implied by its biological functions, the issues not properly investigated till yet. One could have been its position at the end of the genome and beyond the poly-U tract, making the X-tail prone to nuclease degradation. Moreover, due to its functions as a 3′-element it could have been associated strongly with viral or cellular proteins. X-tail-based viral loads were highly concordant with results from bDNA testing. bDNA was chosen as a gold standard because it uses multiple probes along the 5′-NCR and initial core region and has proven to be the most robust viral load test compared to other assays (Drexler et al., 2009; Caliendo et al., 2006; Elbeik et al., 2004; Tuaillon et al., 2007).

The evolution of HCV is a highly dynamic process. It occurs both through multiple processes of adaptive selection that drive sequence change (such as those resulting from the host immune response and potentially from antiviral treatment) and through drift, in which phenotypically neutral sequence changes accumulate over time without altering the phenotype or behaviour of the virus. However, despite its potential to change rapidly, the Longer-term evolution of HCV appears to be remarkably conservative. Whilst the differences in treatment response between genotypes are important clinically, there has been little fundamental change in the relationship between HCV genotypes and their human hosts (such as their ability to persist and transmit) over the extremely long periods over which they have probably evolved. HCV thus appears to have successfully filled an extremely specific ecological niche in human populations. Knowing more about the intimate host–parasite relationship that balances innate and acquired immune-defence mechanisms in the host with the development of complex evasion mechanisms in the virus is the key to understanding its pathogenesis and for developing future treatment intervention strategies (Araujo et al., 2008; Simmonds, 2004).

Regions of the genome such as NS5B have been used frequently for epidemiological reconstruction, other parts of the genome, such as the 'hypervariable' regions (HVRs) of E2 and NS5A; show much greater sequence variability and much more rapid amino acid sequence change over time. One of the reasons of this variability may be specific selection, Darwinian mechanisms operating on the virus that are associated with immune escape, like the HVR in E2 may be a target for neutralizing antibody and persistence might require continuous virus sequence change to evade B-cell responses (Farci et al., 2000; Kantzanou et al., 2003). Many of the amino acid polymorphisms that are observed in HCV are driven sequentially by selection from different major histocompatibility complex class I or II alleles or by antibody recognition that is encountered during a virus's passage through human populations. Indeed, direct evidence for the occurrence of immune-selected changes in cytotoxic T-lymphocyte (CTL) epitopes has been obtained on experimental infection of chimpanzees (Erickson et al., 2001). Sequence change was slower in individuals with defects in T- or B-cell immunity, also supports this hypothesis (Booth et al., 1998), interpreted as indicating

reduced immune selection on CTL or B-cell epitopes. Indeed, recovery from infection is linked with strong and sustained CTL responses during the time of primary infection, a time where there is evidence for specific changes in CTL epitopes and accelerated sequence change in the coding sequence in those who become chronic carriers (Lechner et al., 2000; Thimme et al., 2001; Cantaloube et al., 2003; Sheridan et al., 2004). Immune selection may also underlie the high degree of sequence variability between and within genotypes in the NS5A region and therefore lead to differences in the ability of various variants and genotypes to evade intracellular defenses.

## QUISISPECIES

Continued fixation of nucleotide changes over time and the development of variable degrees of sequence diversity within the replicating population at a given time frame, are the processes leading to neutral and adaptive evolution of HCV during the course of chronic infection within an individual. Sequence diversity is a continuous process during virus replication, as RNA copying by the virally encoded RNA polymerase (NS5B) is error-prone and the replicating population is so large. For example, ongoing error rates of between 1 in 10 000 and 1 in 100 000 bp copied, which are typically found for RNA polymerases combined with a rate of virus production of up to 1012 virions per day, would produce a highly genetically diverse population of variants, containing mutants that differed at every nucleotide position and every combination of paired differences from the population mean or consensus (Drake et al., 1998; Neumann et al., 1998).

Even though the consensus sequence may be close to the fitness peak at any one time, the existence of a large and diverse population would allow rapid, adaptive, Darwinian changes in response to changes in the replication environment. Further the selection against viruses with specific T- or B-cell epitopes might take the form of evolving immune responses; that might also confer resistance to antiviral therapies. The rapid and reproducible independent appearance of specific amino acid changes that are associated with the acquisition of HIV-1 resistance to reverse transcriptase and protease inhibitors is a dramatic demonstration of Darwinian evolution of the 'quasispecies'. In the future, this may be reproduced in HCV infections that are treated with the new generation of protease and RNA polymerase inhibitors (Lamarre et al., 2003; Pause et al., 2003; Trozzi et al., 2003; Lu et al., 2004; Sarisky, 2004).

## RECOMBINATION

Recombination occurs in many families of RNA viruses, its occurrence requiring both epidemiological opportunity and biological compatibility. In positive-stranded RNA viruses, recombination generally occurs through a process of template-switching during RNA genomic replication. To detect such occurrences, a single cell must be infected with two or more genetically identifiable variants of the virus. *In vivo*, this requires both coinfection of the same individual with more than one such variant and substantial overlap in their geographical distributions, in order to enable recombinant forms to be detected (Simmonds, 2004).

The genotype, epidemiology and natural history of infection with HCV clearly fulfils both of these criteria. A wide range of genotypes circulate in the main risk groups for HCV in Western countries, including 1a and 3a in IDUs and 1b, 2a-2c and 4a throughout the Mediterranean area. In these areas, infection is often characterized by multiple exposures around the time of primary infection, such as frequently repeated needle-sharing with several infected individuals over short time-intervals in the case of IDUs and the contamination of blood products, such as factor VIII clotting factor concentrates, with multiple HCV-positive plasma units. Indeed, even ongoing, chronic HCV infection does not protect from reinfection in experimentally challenged chimpanzees (Farci et al., 1992) or in HCV-contaminated blood or blood-product recipients, such as thalassaemics and hemophiliacs (Lai et al., 1994). The viability and pathogenicity of inter- and intra-genotype recombinants are more difficult to assess and are likely to vary considerably between virus families.

There is little experimental information on the potential viability of inter- or intra-genotype recombinants of HCV, although it has recently been observed that most combinations of genotype 1a and 1b sequences in the non-structural region of the genome fail to generate a viable replicon (Gates et al., 2004), implying the existence of incompatibilities between variants that show approximately 20% sequence divergence. Despite these *in vitro* observations, recombinant forms of HCV have been observed in nature, including a variant formed from structural genes of genotype 2 and non-structural genes from genotype 1b that was found in infected IDUs in St Petersburg, Russia (Kalinina et al., 2002, 2004) and a possible 1a/1b recombinant in Peru (Colina et al., 2004). Despite the number of studies that have been carried out to investigate this issue, the true frequency of recombination of HCV may have been considerably underestimated. Recombination would not be detected easily between variants of the same subtype (such as between two infecting genotype 1a strains in an IDU). Similarly, it would be difficult to document inter-subtype recombinants where HCV is highly diverse, such as within genotype 2 in western Africa, because in such regions, a full catalogue of sequence variants is lacking within which to observe recombination events (Simmonds, 2004).

## CONCLUSION

The emergence of widespread recombination would place a

considerable limitation on the use of genotyping assays presently in use that are based on single genome regions, such as the UTR or core gene. As more recombinant viruses emerge in the coming years as a result of increasing geographical overlap in genotype distributions, this would cause major problems in the interpretation of genotyping assay results. For recombinant viruses, only those assays that genotyped samples in regions of the genome that determined IFN susceptibility would be able to predict treatment response, which is one of the main applications of genotyping assays. Secondly, reconsideration of hepatitis C virus classification is required for the better understanding of the treatment response.

**REFERENCES**


Allain JP, Sarkodie F, Boateng P, Asenso K, Kyeremateng E, Owusu-Ofori S (2008). A pool of repeat blood donors can be generated with little expense to the blood center in sub-Saharan Africa. Transfus. 48: 735-741.

Araujo F, Sonoda IV, Rodrigues NB, Teixeira R, Redondo R, Oliveira GC (2008). Genetic variability in the 5` UTR and NS5A regions of hepatitis C virus RNA isolated from non-responding and responding patients with chronic HCV genotype 1 infection. Mem. Inst Oswaldo Cruz, Rio de Janeiro. 103: 611-614.

Booth JC, Kumar U, Webster D, Monjardino J, Thomas HC (1998). Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogamma globulinemic patients. Hepatology, 27: 223-227.

Buendia P, Cadwallader B, DeGruttola V (2009). A phylogenetic and Markov model approach for the reconstruction of mutational pathways of drug resistance. Bioinfo. Advance Access published online.

Bukh J (1995). Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin Liv. Dis. 15: 41-63.

Bukh J, Purcell RH, Miller RH (1993). At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc. Nat. Acad. Sci. USA, 90: 8234-8238.

Caliendo AM, Valsamakis A, Zhou Y, Yen-Lieberman B, Andersen J, Young S, Ferreira-Gonzalez A, Tsongalis GJ, Pyles R, Bremer WJ, Lurain NS (2006). Multilaboratory comparison of hepatitis C virus viral load assays. J. Clin. Microbiol. 44: 1726-1732.

Cantaloube JF, Biagini P, Attoui H, Gallian P, De Micco P, De Lamballerie X (2003). Evolution of hepatitis C virus in blood donors and their respective recipients. J. Gen. Virol. 84: 441-446.

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

Cristina J (2005). Genetic diversity and evolution of hepatitis C virus in the Latin American region. J. Clin. Virol. 34: S1-S7.

Cristina J, Mukomolov S, Colina R, Kalinina O, Garcia L, Khan B, Mogdasy C, Karayiannis P (2002). Hepatitis C virus phylogeny: A useful clinical tool. Acta Virol. 46: 179-182.

Dienstag JL, Isselbacher JK (2005). Acute Viral Hepatitis. In: Harrison's Principles of Internal Medicine. 16th Edition. McGraw-Hill Co. N.Y. pp. 1801-1805.

Dixit S, Quan V, Martin P, Larson D, Brezina M, Dinello R, Sra K, Lau YN, Chien D, Kolberg J, Tagger A, Davis G, Polito A, Gitnick G (1995). Evaluation of a Novel Serotyping System for Hepatitis C Virus: Strong Correlation with Standard Genotyping Methodologies. J. Clin. Virol. 33: 2978-2983.

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998). Rates of spontaneous mutation. Genetics, 148: 1667-1686.

Drexler FJ, Kupfer B, Petersen N, Grotto RMT, Rodrigues SMC, Grywna K, Panning M, Annan A, Faria Silva G, Douglas J, Koay ESC, Smuts H, Netto EM, Simmonds P, Pardini MC, Roth WK, Drosten C (2009). A Novel Diagnostic Target in the Hepatitis C Virus Genome. Feb. Iss. PloS. Med. 6: 0210-0220.

Elbeik T, Surtihadi J, Destree M, Gorlin J, Holodniy M, Jortani SA, Kuramoto K, Ng V, Valdes Jr. R, Valsamakis A, Terrault NA (2004). Multicenter evaluation of the performance characteristics of the Bayer Versant HCV RNA 3.0 assay (bDNA). J. Clin. Microbiol. 42: 563-569.

Erickson A, Kimura L, Igarashi Y, Igarashi S, Eichelberger J, Houghton M, Sidney J, McKinney D, Sette A, Hughes AL, Walker CM (2001). The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic *T. lymphocytes*. Immunity, 15: 883-895.

Farci P, Alter HJ, Govindarajan S, Wong DC, Engle R, Lesniewski RR, Mushahwar IK, Desai SM, Miller RH, Ogata N, Purcell H (1992). Lack of protective immunity against reinfection with hepatitis C virus. Science, 258: 135-140.

Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzera A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ (2000). The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science, 288: 339-344.

Gates AT, Sarisky RT, Gu B (2004). Sequence requirements for the development of a chimeric HCV replicon system. Virus Res. 100: 213-222.

Germer JJ, Zein NN (2001). Advances in the molecular diagnosis of hepatitis C and their clinical implications. Mayo. Clin. Proc. 76: 911-920.

Henry P, Zhiying H, Austin LH, William B (2004). Perinatal Transmission and Viral Evolution of Hepatitis C Virus Quasispecies in Infants Coinfected With HIV. J. A.I.D.S. 36: 890-899.

Honardoost M, Sabahi F, Bavil-Olyaee SA, Behzadian F, Merat S, Malekzdeh R (2008). Interferon resistence of hepatitis C Virus genotype 1a/1b relationship to structural E2 gene. Quasispecies mutations. Iran J. Biotechnol. 6: 36-44.

Kalinina O, Norder H, Magnius LO (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.

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

Kantzanou M, Lucas M, Barnes E, Komatsu H, Dusheiko G, Ward S, Harcourt G, Klenerman P (2003). Viral escape and T cell exhaustion in hepatitis C virus infection analysed using class I peptide tetramers. Immunol. Lett. 85: 165-171.

Kavita SL, Jha JA, Shrotri SP, Tandon BN, Prasad VGM, Arankalle VA (2003). Comparison of Hepatitis C Virus Genotyping by 5' Noncoding Region- and Core-Based Reverse Transcriptase PCR Assay with Sequencing and Use of the Assay for Determining Subtype Distribution in India. J. Clin. Microbiol. 41: 5240-5244.

Kronke J, Kittler R, Buchholz F, Windisch MP, Pietschmann T, Bartenschlager R, Frese M (2004). Alternative Approaches for Efficient Inhibition of Hepatitis C Virus RNA Replication by Small Interfering RNAs. J. Virol. 78: 3436-3446.

Kuntz E, Kuntz HD (2006). Hepatology principles and practices. History, Morphology, Biochemistry, Diagnostics, Clinical therapy. Springer, 2nd Edition. p. 413.

Laar T, Pybus O, Bruisten S, Brown D, Nelson M, Bhagani S, Vogel M, Baumgarten A, Chaix ML, Fisher M, Gőtz H, Matthews GV, Neifer S, White P, Rawlinson W, Pol S, Rockstroh J, Coutinho R, Dore GJ, Geoffrey M, Dusheiko GM, Danta M (2009). Evidence of a Large, International Network of HCV Transmission in HIV-Positive Men Who Have Sex With Men. Gastroenterology, 136: 1609-1617.

Lai ME, Mazzoleni AP, Balestriesi A, Argiolu F, De Virgilis S, Cao A, Purcell RH, Farci P (1994). Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. Lancet, 343: 388-390.

Lamarre D, Anderson PC, Bailey M, Beaulieu P, Bolger G, Bonneau P, Bös M, Cameron DR, Cartier M, Cordingley MG, Faucher AM, Goudreau N, Kawai SH, Kukolj G, Lagacé L, LaPlante SR, Narjes H, Poupart MA, Rancourt J, Sentjens RE, St George R, Simoneau B, Steinmann G, Thibeault D, Tsantrizos YS, Weldon SM, Yong CL, Llinàs-Brunet M (2003). An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. Nature, 426: 186-189.

Lechner F, Wong DKH, Dunbar PR, Chapman R, Chung RT, Dohrenwend P, Robbins G, Phillips R, Klenerman P, Walker BD (2000). Analysis of successful immune responses in persons infected with hepatitis C virus. J. Exp. Med. 191: 1499-1512.

Lonardo A, Adinolfi LE, Loria P, Carulli N, Ruggiero G, Day CP (2004). Steatosis and Hepatitis C virus: Mechanisms and significance for hepatic and extrahepatic disease. Gastroenterology, 126: 586-597.

Lu L, Pilot-Matias TJ, Stewart KD, Randolph JT, Pithawalla R, He W, Huang PP, Klein LL, Mo H, Molla A (2004). Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor in vitro. Antimicrob. Agents Chemother. 48: 2260-2266.

Markov PV, Pepin J, Frost E, Deslandes S, Labbé AC, Pybus OG (2009). Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. J. Gen. Virol. 90: 2086-2096.

Mizokami M, Tanaka Y (2005). Tracing the Evolution of Hepatitis C Virus in the United States, Japan and Egypt By using the molecular clock. Clin. Gastroenterol. Hepatol. 3: S82-S85.

Moratorio G, Martínez M, Gutiérrez MF, González K, Colina R, López-Tort F, López L, Recarey R, Schijman AG, Moreno MP, García-Aguirre L, Manascero AR, Cristina J (2007). Evolution of naturally occurring 5'non-coding region variants of Hepatitis C virus in human populations of the South American region. Virology, 79: 1-12.

Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS (1998). Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-a therapy. Science, 282: 103-107.

Ohno O, Mizokami M, Wu RR, Saleh MG, Ohba K, Orito E, Mukaide M, Williams R, Lau JY (1997). New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. J. Clin. Microbiol. 35: 201-207.

Okamoto H, Tokita H, Sakamoto M, Horikita M, Kojima M, Iizuka H, Mishiro S (1993). Characterization of the genomic sequence of type V (or 3a) hepatitis C virus isolates and PCR primers for specific detection. J. Gen. Virol. 74: 2385-2390.

Pause A, Kukolj G, Bailey M, Brault M, Dô F, Halmos T, Lagacé L, Maurice R, Marquis M, McKercher G, Pellerin C, Pilote L, Thibeault D, Lamarre D (2003). An NS3 serine protease inhibitor abrogates replication of subgenomic hepatitis C virus RNA. J. Biol. Chem. 278: 20374-20380.

Pawlotsky JM (2003). Hepatitis C virus genetic variability: pathogenic and clinical implications. Clin. Liver Dis. 7: 45-66.

Pawlotsky JM (2009). Review: Therapeutic implications of hepatitis C virus resistance to antiviral drugs. Ther. Adv. Gastroenterol. 2: 205-219.

Roque-Afonso AM (2005). Compartmentalization of hepatitis C virus genotypes between plasma and peripheral blood mononuclear cells. Am. J. Virol. 79: 6349-6357.

Sarisky RT (2004). Non-nucleoside inhibitors of the HCV polymerase. J. Anti. Microb. Chemoth. 54: 14-16.

Sheridan I, Pybus OG, Holmes EC, Klenerman P (2004). High-resolution phylogenetic analysis of hepatitis C virus adaptation and its relationship to disease progression. J. Virol. 78: 3447-3454.

Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, Brouwer JT, Chan SW, Chayama K, Chen DS, Choo QL, Colombo M, Cuypers HTM, Date T, Dusheiko GM, Esteban JI, Fay O, Hadziyannis JS, Han J, Hatzakis A, Holmes EC, Hotta H, Houghton M, Irvine B, Kohara M, Kolberg JA, Kuo G, Johnson J, Lau YN, Lelie PN, Maertens G, McOmish F, Miyamura T, Mizokami M, Nomoto A, Prince AM, Reesink HW, Rice C, Roggendorf M, Schalm SW, Shikata T, Shimotohno K, Stuyver L, Trépo C, Weiner A, Yap PL, Urdea MS (1993). A proposed system for the nomenclature of hepatitis C viral genotypes. J. Gen. Virol. 74: 661-668.

Simmonds P (2004). Genetic diversity and evolution of hepatitis C virus-15 years on. J Gen. Virol. 85: 3173-3188.

Stuyver L, Rossau R, Wyseur A, Duhamel M, Vanderborght B, Van Heuverswyn H, Maertens G (1993). Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. J. Gen. Virol. 74: 1093-1102.

Stuyver L, Vanarnhem W, Wyseur A, Hernandez F, Delaporte Eo, Haertens G (1994). Classification of hepatitis C viruses based on phylogenetic analysis of the envelope I and nonstructural 5b regions and identification of five additional subtypes. Proc. Nat. Acad. Sci. USA, 91: 10134-10138.

Tanaka Y, Kurbanov F, Mano S, Orito E, Vargas V, Esteban JI, Yuen MF, Lai CL, Kramvis A, Kew MC, Smuts HE, Netesov SV, Alter HJ, Mizokami M (2006). Molecular tracing of the Global Hepatitis C Virus Epidemic Predicts Regional patterns of Hepatocellular Carcinoma Mortality. Gastroenterol. 130: 703-714.

Tatjana T, Helene N, Valentine T, Tonu K, Jelena S, Madis I, Karen O, kaije P, Ludmilla P, Lars MO (2007). Genetic characterization of hepatitis c virus strains in Estonia: fluctuations in the predominating subtype with time. J. Med. Virol. 79: 374-382.

Thimme R, Oldach D, Chang KM, Steiger C, Ray SC, Chisari FV (2001). Determinants of viral clearance and persistence during acute hepatitis C virus infection. J. Exp. Med. 194: 1395-1406.

Tuaillon E, Mondain AM, Ottomani L, Roudiere L, Perney P, Caliendo AM, Valsamakis A, Zhou Y, Yen-Lieberman B, Andersen J, Young S, Ferreira-Gonzalez A, Tsongalis, GJ, Pyles R, Bremer JW, Lurain NS (2007). Impact of hepatitis C virus (HCV) genotypes on quantification of HCV RNA in serum by COBAS AmpliPrep/COBAS TaqMan HCV test, Abbott HCV realtime assay and VERSANT HCV RNA assay. J. Clin. Microbiol. 45: 3077-3081.

Tuplin A, Evans DJ, Simmonds P (2004). Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. J. Gen. Virol. 85: 3037-3047.

Varaklioti A, Vassilaki N, Georgopoulou U, Mavromara P (2002). Alternate translation occurs within the core coding region of the hepatitis C viral genome. J. Biol. Chem. 277: 17713-17721.

Walewski JL, Keller TR, Stump DD, Branch AD (2001). Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. RNA, 7: 710-721.

Wesley A, Alter JM (2000). Epidemiology of hepatitis C: Geographic differences and temporal trends. Semin. Liver Dis. 20: 1-16.

Weiner AJ, Thaler MM, Crawford K, Ching K, Kansopon J, Chien DY, Hall JE, Hu F, Houghton M (1993). A unique, predominant hepatitis C virus variant found in an infant born to a mother with multiple variants. J. Virol. 67: 4365-4368.

Xu Z, Choi J, Yen TSB, Lu W, Strohecker A, Govindarajan S, Chien D, Selby MJ, Ou J (2001). Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. EMBO J. 20: 3840-3848.

Yuan HJ, Jain M, Snow KK, Gale Jr. M, Lee WM (2009). Evolution of hepatitis C virus NS5A region in breakthrough patients during pegylated interferon and ribavirin therapy. J. Viral. Hep. 1-9.

Zein NN (2000). Clinical Significance of Hepatitis C Virus Genotypes. Clin. Microbiol. Rev. 13: 223-235.