Introduction

Hepatitis C virus (HCV) is an enveloped single-stranded RNA virus that contains approximately 9600 nucleotides in the genome, and is one of the major causes of chronic hepatitis worldwide [1,2,3]. HCV isolates exist as a group of heterogeneous but genetically distinct variants known as quasispecies. Sequence and phylogenetic analysis of the various HCV genomic regions such as the 5’untranslated region (5’UTR), envelope (E1), core, and non-structural 5B (NS5B) led to the identification of six major genotypes, and more than 90 closely related subtypes within the genotypes [4,5].

Several studies have shown that the prevalence of subtypes varies among different geographical regions of the world and that severity of liver disease as well as sensitivity to antiviral treatment varies with different subtypes. In the Philippines, HCV-1b (50%) was the most common subtype among blood donors, followed by HCV-1a (36%). Among inmates, subtype 1a (68%) was the most common, followed by subtype 1b (11%). It has been reported that the positive rate for anti-HCV was 2.2% among blood donors and 4.6% among inmates in the Philippines [6,7]. Recently, it was reported that among intravenous drug users in Metro Cebu 83% were positive for hepatitis C. The predominant HCV subtype was 1a. However, subtype 1b was found to occur only in Metro Manila [8].

The HCV genome contains a highly conserved region, the 5’UTR, which is often targeted to detect major genotypes. However, this region cannot accurately discriminate genotypic
subtypes, because it is too conserved or not heterogenous enough [9,10]. It has been reported that only a single base change at position -99 (adenine to guanine) distinguishes between subtypes 1a and 1b in the 5'UTR sequence. Recent investigations have found that methods that rely on the 5'UTR are unable to differentiate subtypes 1a from 1b in approximately 10% of the cases. In addition, it seems that some of the specific sequence motifs identified in HCV 5'UTR are no longer found to be conserved. The guanine (G) residue, for example, at position -99 is found in subtype 1a viruses and the adenine (A) residue at position -99 in subtype 1b viruses. Thus, may lead to misclassifications when 5'UTR-based typing techniques are used [11].

Alternative genomic regions have been proposed for use in HCV typing. The non-structural 5B (NS5B) for example, is useful because it is sufficiently variable which may identify both genotypes and subtypes. Sequence and phylogenetic analysis of variable genomic regions such as the NS5B has been recommended for HCV genotyping and subtyping [12,13,14,15]. In this study, we subtyped HCV-1 samples by direct nucleotide sequencing with nested RT-PCR products derived from the 5'UTR and the NS5B region.

Materials and methods

Samples and patients

Thirty blood samples collected from May 2005 to December 2008 from patients infected with HCV-1 previously confirmed by PCR-RFLP and clinically diagnosed with chronic hepatitis C at St. Luke’s Medical Center, Philippines were analyzed. There were 17 males and 13 females with ages ranging from 32 to 76 years old. Patients are excluded if they have other causes of liver diseases such as autoimmune hepatitis and history of alcoholism, or reactive to hepatitis B surface antigen and antibody to hepatitis B core antigen.

Viral RNA extraction and cDNA synthesis

The viral nucleic acid from HCV-infected patient plasma was extracted from peripheral blood using the QIAamp® Viral RNA Mini kit from Qiagen according to manufacturer’s instructions. Reverse transcription was carried out from 10 µl of viral RNA extract using the SuperScript™ III reverse transcriptase (Invitrogen™).

5'UTR nested PCR amplification

Two sets of oligonucleotide primers were used to amplify a 251-bp fragment of the 5’UTR. The first round PCR reaction was performed as follows: 4 µl of the cDNA was added to a 46 µl of PCR mixture containing 0.5X Phusion® HF buffer, 1.0 pmol of outersense and outer antisense primer, 0.10 mM dNTP, 0.01 U Phusion® DNA polymerase, and DEPC-treated water. Amplification of the 5’UTR was carried out at 94°C for 1 minute followed by 25 cycles of denaturation at 94°C for 25 seconds, annealing at 50°C for 40 seconds, extension at 72°C for 60 seconds, and final elongation at 72°C for 10 minutes. The first PCR product was used as a template for the second PCR amplification. One microliter of the reaction mixture was transferred to a second tube containing the same medium but with the inner pair of sense and antisense primers (Table 1).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polarity</th>
<th>Nucleotide Sequence (5’ to 3’)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'UTR</td>
<td>Outersense</td>
<td>CTGTGAGGAACACTGCTTT</td>
<td>Chan et al., 1992</td>
</tr>
<tr>
<td></td>
<td>Outer antisense</td>
<td>ATACTCGAGGTGCACGGTCTACGAGACCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Innersense</td>
<td>TTCACGCAGAAAGCGTCTAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inner antisense</td>
<td>CACTCTCGAGCCCTATCAGGCAGT</td>
<td></td>
</tr>
<tr>
<td>NS5B</td>
<td>Outersense</td>
<td>TGGGTTTCTCGTATGATACCC</td>
<td>Chen et al., 2002</td>
</tr>
<tr>
<td></td>
<td>Outer antisense</td>
<td>CCTGGTCATAGCGCTCCGTGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Innersense</td>
<td>GATACCGCTGGTTTACCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inner antisense</td>
<td>CCTCCGTGAAGGTCTCAG</td>
<td></td>
</tr>
</tbody>
</table>
5'UTR and NS5B nucleotide sequences in hepatitis C virus subtyping

Table 2. Concordance of HCV-1 subtyping by 5'UTR and NS5B sequencing

<table>
<thead>
<tr>
<th>NS5B subtyping</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'UTR sequencing</td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>11</td>
</tr>
<tr>
<td>1b</td>
<td>2*</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
</tr>
</tbody>
</table>

The 5'UTR sequencing is horizontally indicated and the NS5B is presented vertically. * Inconsistent subtype assignment.

NS5B nested PCR amplification

The nested PCR was carried out on 1 µl of the first amplicon containing 1X Phusion™ HF buffer, 0.5 pmol of outersense and outer antisense primer, 0.2 mM dNTP, 0.02 U Phusion™ DNA polymerase, and DEPC-treated water with the following thermal profile: initial denaturation at 95°C for 1 minute followed by 40 cycles at 95°C for 20 seconds, 56°C for 30 seconds, 72°C for 1 minute and final extension at 72°C for 10 minutes. The conditions of the second PCR were the same as described above using inner pair of sense and antisense primers (Table 1). The nested PCR product was electrophorized in ethidium bromide-stained 2% agarose gel and then visualized under an ultraviolet transilluminator to identify the desired 367-bp fragments. The NS5B was chosen because this region contains subtype specific motifs and it could be readily amplified from plasma of HCV-infected individuals. However, the use of other coding regions of the HCV genome such as the core and E1 may be helpful to improve the accuracy of HCV-1 subtyping.

DNA purification and direct nucleotide sequencing of the 5'UTR and NS5B

Nested PCR products of 5'UTR and NS5B regions were purified using the GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare) according to the manufacturer’s protocol. Amplicons in the 5'UTR (251-bp) and NS5B (367-bp) regions were sequenced using Big Dye Terminator Cycle Sequencing Ready Reaction Kit and Applied Biosystem 3730xl automated sequencer (Macrogen, Korea). Sequence data were aligned with the consensus sequences of known subtypes using ChromasPro version 1.34 and BioEdit version 7.0 and identity sequences were compared using the basic local alignment search tool (BLAST) program. The GenBank/EMBL/DDBJ accession numbers of HCV sequences used in the analysis were M62321 (HCV-1) for 1a, D90208 (HCV-J) for 1b, AF238485 (MD2a-7) for 2a, D10988 (HC-J8) for 2b, D17763 (NZL1) for 3a, Y11604 (ED43) for 4a, Y13184 (EUH1480) for 5a, and Y12083 (EUHK2) for 6a.

Phylogenetic analysis of the 5'UTR and NS5B

Phylogenetic trees of the 5'UTR and NS5B regions were constructed using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 in accordance to the neighbor-joining (NJ) method. The NJ trees were bootstrapped 1000 times to obtain the final tree. The significance of the group was assumed when bootstrap values were greater than 70%.

Nucleotide sequence accession numbers

The nucleotide sequences reported in this study have been submitted to GenBank. It can be retrieved under GenBank accession numbers GQ844690 to GQ844700 and GQ866987 to GQ8667012.

Results

Sequence analysis of the NS5B showed that 13 (43%) of the samples belonged to subtype 1a and 17 (57%) of the samples belonged to subtype 1b. The most predominant subtype was 1b by NS5B sequencing. Using NS5B as the reference method, 5'UTR sequencing correctly identified subtype 1a in 11/13 (85%) and subtype 1b in 13/17 (76%) of the samples. Eleven of the 15 samples were correctly typed as HCV-1a by 5'UTR sequencing, while the remaining 4 samples identified as 1a by 5'UTR sequencing were confirmed as 1b by NS5B sequencing. The predictive value of 5'UTR sequencing to subtype 1a was 73%. On the other hand, 13 of the 15
samples were correctly typed as HCV-1b by 5′UTR sequencing, while the remaining 2 samples giving discrepant result identified as 1b by 5′UTR sequencing was confirmed as 1a by NS5B sequencing. The predictive value of 5′UTR sequencing to subtype 1b was 87%. Overall, of the 30 HCV-1 samples, 24 (80%) were concordantly subtyped by 5′UTR sequencing (Table 2).

The key amino acids associated with 1a viruses were aspartic acid (D) at position 2665, glutamine (Q) at position 2667, valine (V) at position 2670, lysine (K) at position 2673, arginine (R) at position 2689, glutamic acid (E) at position 2691, isoleucine (I) at position 2716, arginine (R) at position 2719, and glycine (G) at position 2726. On the other hand, the key amino acids associated with 1b viruses were alanine (A) at position 2665, glutamic acid (E) at position 2667, glutamine (Q) at position 2670, arginine (R) at position 2673, lysine (K) at position 2689, glutamine (Q) at position 2691, leucine (L) at position 2716, threonine (T) at position 2719, and lysine (K) at position 2726 (Figure 1, 2).

Discussion

Many molecular biology techniques have been developed and have been proven to be useful
5’UTR and NS5B nucleotide sequences in hepatitis C virus subtyping

Figure 3. Phylogenetic analysis of partial NS5B sequences of 30 HCV-1 samples. HCV prototype sequences from GenBank/EMBL/DDBJ were included. The evolutionary history was inferred using the Neighbor-Joining (NJ) method. The numbers at the nodes represent the percent bootstrap support for 1000 replicates. Only values over 70% are shown. Bar at the base of the tree shows the genetic divergence. Phylogenetic analysis was conducted in MEGA4.

for laboratories. DNA sequencing represents one of these useful tools. Results showed that direct nucleotide sequencing of the 5’UTR did not resolve completely all existing HCV-1 subtypes in 20% of the cases. The discrepant results have been previously reported in other studies. In this regard, it is important to perform nucleic acid sequencing of variable genomic regions such as the NS5B, which exhibit a greater degree of nucleotide heterogeneity within a given genotype.

Chen et al., demonstrated concordant results between 5’UTR and NS5B sequencing in 80% of cases [9]. In another study, Tamalet et al., reported 70% concordance between the two methods [16]. In this study, the overall concordance between 5’UTR and NS5B sequencing was 80%. The probable reason for this is most likely due to the same set of primers used in this study and with Chen et al., [9]. In addition, NS5B nested PCR amplification was carried out in this study instead of single-step PCR amplification as performed by Tamalet et al.,[16]. Phylogenetic analysis of the NS5B showed that all of subtype 1b viruses grouped together and all of subtype 1a viruses clustered together supported by bootstrap value of 99% (Figure 3). The subtype assignments of these samples were confirmed by basic local alignment search tool program (BLAST) searches. Clearly, only methods based on the use of more variable regions such as the NS5B can be relied upon for accurate discrimination of subtypes. In addition, sequence and phylogenetic analysis of the NS5B region could be the first step in molecular epidemiological studies designed to identify the route of HCV transmission as described by Sandres-Saune et al.,[17].

Studies have shown that the 5’UTR, is 92-98% conserved and contains subtype specific sequence motifs [18]. Consistent with other reports, we have demonstrated that subtype 1a had adenine (A) residue at position -99 and subtype 1b had guanine (G) residue at position -99.
5′UTR and NS5B nucleotide sequences in hepatitis C virus subtyping

(Figure 4, 5). In addition, the guanine (G) residue was found to occur in all of subtype 1b except for 1 sample (#814) which was typed as 1a by PCR-RFLP but had a guanine (G) residue at position -99. This finding was not surprising and could be explained by an A/G polymorphism that exist at position -99 or the presence of quasispecies in the infected patient. It has been shown that HCV-1a and 1b viruses have been frequently mistyped, mostly because subtyping in the 5′UTR is based on a single-base polymorphism at position -99. Furthermore, nucleotide sequencing of the 5′UTR alone cannot accurately differentiate HCV-1 subtypes because it is too conserved for discrimination of subtypes. It has been suggested that a rate of mistyping by 16% may result in an equilibration of the observed shift in subtype prevalence [19]. Thus, sequence analysis based on the nucleotide signature motifs should be interpreted with caution.

Figure 4. Alignment of partial HCV-1a 5′UTR sequences. At the top, HCV-1a reference sequence (GenBank accession number M62321) is given. At the left, the sample identification numbers are shown. The key nucleotide associated with 1a is boxed. At the right, the subtyping results observed by the 5′UTR sequencing are shown.

D90208 (1b)  TTGGATCAACCGCTCAATGCCTGGAGATTTGGGCCGTGCCCGCCGAGACTGCTA  Subtype
443  -------------------------------  1b
544  G---------T---------------  1b
551  G---------T---------------  1b
620  G---------T---------------  1b
622  G---------T---------------  1b
645  G---------T---------------  1b
691  G---------T---------------  1b
701  G---------T---------------  1b
743  G---------T---------------  1b
745  G---------T---------------  1b
751  G---------T---------------  1b
785  G---------T---------------  1b
814  G---------T---------------  1b
818  G---------T---------------  1b

Figure 5. Alignment of partial HCV-1b 5′UTR sequences. At the top, HCV-1b reference sequence (GenBank accession number D90208) is given. At the left, the sample identification numbers are shown. The key nucleotide associated with 1b is boxed. At the right, the subtyping results observed by the 5′UTR sequencing are shown.
5'UTR and NS5B nucleotide sequences in hepatitis C virus subtyping

because this could lead to mistyping.

Phylogenetic analysis of the 5'UTR showed that not all of subtype 1b viruses consistently grouped together and not all of subtype 1a viruses clustered together (Figure 6). The results suggested that sequence and phylogenetic analysis of the 5'UTR cannot accurately differentiate HCV-1a and 1b subtypes in (6/30) 20% of cases as compared with the analysis based on NS5B region.

Results of HCV-1 subtyping obtained in this study show that reliable distinction between 1a and 1b viruses can be achieved when the decision is not exclusively based on sequence information derived from the highly conserved 5'UTR of the HCV genome. Despite mounting evidence that analysis of the 5'UTR does not accurately discriminate HCV-1 subtypes, many laboratories still use the 5'UTR analysis for identification of HCV-1 subtypes [9]. Most commercially available HCV typing assays target the 5'UTR, since this region is the most highly conserved and suited for amplification methods. However, it has only a limited subtyping accuracy as shown by this study and is in agreement with the studies of Verbeeck et al., [20].

The choice of the genome region to be analyzed in the identification of HCV-1 subtypes is crucial. First, this region must contain subtype specific motifs, which represent the diversity of the entire viral genome. Second, sufficient conservation within the subtype is important for the development of primers, so that it can be readily detected by assays based on nucleic acid amplification. Third, sufficient variability is needed to allow discrimination between subtypes [17]. Although the preferred region for subtyping is the NS5B, it is not always possible to amplify this region because of primer-target mismatch within the highly variable NS5B sequence. Failure in sequencing the NS5B gene has been reported by some investigators despite successive attempts to amplify cDNA with the NS5B primers [16]. Despite this inability to possibly amplify HCV RNA samples, NS5B sequence analysis was used in this study as a reference method for HCV subtyping based on consensus proposals. It is worthwhile to mention that all 30 samples analyzed were successfully amplified and sequenced in this study. A possible explanation for this was a higher viral load in the patients’ plasma tested.

Since therapeutic decisions for chronic HCV-infected patients are made on the basis of genotype, it is important that the genotype and the subtype be accurately determined. Furthermore, accurate identification of both genotypes and subtypes will enable the proper choice of new antiviral compounds which are likely to show distinct activities against isolates belonging to different subtypes of HCV-1.

Intergenotypic and intragenotypic recombinants have been reported from isolates in St. Peters-
5’UTR and NS5B nucleotide sequences in hepatitis C virus subtyping

burg (2k/1b), the Philippines (2b/1b), Vietnam (2/6), Ireland (2k/1b), France (2/5) and in Peru (1b/1a). Recombination is a cause of genetic diversity in some families of RNA viruses including HCV, which has important implications to epidemiologic studies, pathogenesis, diagnosis and treatment of HCV infection [7,21,22,23, 24,25]. Although recombinants of HCV have been reported rarely, this study showed that sequence and phylogenetic analysis of partial 5’UTR and NS5B regions did not provide evidence for recombinant forms. However, further sequence analysis of the full-length regions of the 5’UTR and NS5B is recommended to entirely rule out this phenomenon. This serves as a limitation of this study.

Taken together, these findings have major implications for molecular epidemiologic studies such as monitoring the distribution of the virus strains and identifying risk factors for transmission, as well as investigating clinical differences between patients infected with HCV-1a or 1b. Indeed, several studies reported differences in treatment outcomes between patients infected with 1a and 1b and have used the NS5B region to identify HCV subtypes. Recently, it has been reported that among patients with chronic hepatitis C treated with pegylated interferon and ribavirin for 48 weeks, the overall rate of sustained virologic response (SVR) was 37.8%. They also noted that HCV-1a was associated with lower treatment response than subtype 1b [27].

From a clinical point of view, patients infected with subtype 1b have been found to have more advanced liver disease and higher viral load, which may suggest increased pathogenicity. Thus, accurate subtyping may later on become an indispensable tool in the management of chronic hepatitis C.

Our data should be taken into account when future HCV vaccines are developed. The design of an effective vaccine must consider specific genotypes and subtypes in order to achieve broad protection because of the diversity of HCV variants.

Acknowledgement

This study was supported by a research grant to the St. Luke’s Liver Diseases Study Group (Project No. 06-013) through the Research and Biotechnology Division, St. Luke’s Medical Center Quezon City, Philippines.

Please address correspondence to: Michael O. Baclig, PhD, Research and Biotechnology Division, St. Luke’s Medical Center 279 E. Rodriguez Sr. Blvd. Quezon City, 1102 Philippines. Tel: (632) 726-0467, E-mail: mobaclig@stluke.com.ph

References

[9] Chen Z, Weck K. Hepatitis C virus genotyping: Interrogation of the 5’translated region can-


