Genetic diversity in hepatitis C virus in Egypt and possible association with hepatocellular carcinoma

Mohamed Abdel-Hamid,1,2,3 Mai El-Daly,1,4 Vilma Molnegren,5 Sherif El-Kafrawy,1,4 Sohair Abdel-Latif,6 Gamal Esmat,7 G. Thomas Strickland,2 Christopher Loffredo,8 Jan Albert9 and Anders Widell5

Correspondence
Anders Widell
Anders.Widell@med.lu.se

1Viral Hepatitis Research Laboratory, National Hepatology and Tropical Medicine Research Institute, Cairo, Egypt
2International Health Division, Epidemiology and Preventive Medicine Department, University of Maryland-Baltimore, Baltimore, MD, USA
3Department of Microbiology, Minia University, Egypt
4National Liver Institute, Menoufia, Egypt
5Department of Medical Microbiology, Malmö University Hospital, Lund University, Malmö, Sweden
6National Cancer Institute, Cairo University, Cairo, Egypt
7Department of Tropical Medicine, Cairo University, Cairo, Egypt
8Lombardi Cancer Centre, Georgetown University, Washington DC, USA
9Department of Virology, Immunology and Vaccinology, Swedish Institute for Infectious Disease Control, and Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Solna, Sweden

Egypt has one of the world’s highest prevalences of hepatitis C virus (HCV) infection, with a majority of genotype 4 infections. To explore the genetic diversity of HCV in Egypt, sera from 131 Egyptians [56 from community studies, 37 chronic hepatitis patients, 28 hepatocellular carcinoma (HCC) patients and 10 patients with non-Hodgkin’s lymphoma] were genotyped by restriction fragment-length polymorphism and phylogenetic analyses of sequences from the mid-core and non-structural 5B regions. The different genotyping methods showed good agreement. The majority of the viruses (83 of 131; 63 %) were of subtype 4a, but five other subtypes within genotype 4 were also observed, as well as three genotype 1b, five genotype 1g and one genotype 3a samples. Interestingly, subtype 4o, which was easily identifiable in all three genomic regions, showed an association with HCC (P=0.017), which merits further investigation.
atypical RFLP patterns, representing different geographical, epidemiological and clinical settings, including malignant complications such as HCC cases. In addition to RFLP, we performed phylogenetic analyses on sequences from the mid-core region (217 nt) and the NS5B region (339 nt). When relevant, we also conducted partial sequencing of the 5'-UTR.

The study included 140 HCV RNA-positive Egyptian serum samples that were collected in 2002 and genotyped by RFLP analysis at the Viral Hepatitis Research Laboratory, Cairo, Egypt. Samples were selected based solely on RFLP results: 35 samples had atypical restriction sites (nontypable), whereas 105 had typical restriction sites (typable) (McOmish et al., 1994). Thus, the age, sex, clinical and epidemiological statuses of the subjects were not considered when the samples were selected. Of these 140 samples, 131 were considered further because they were sequenced successfully in the mid-core region (99 from RFLP-typable samples, including 95 of type 4, three of type 1 and one of type 3, whilst 32 were RFLP non-typable). The geographical origins and clinical characteristics of the 131 samples were diverse: 56 specimens were from a community survey in the southern Nile Valley or from a vertical-transmission study in pregnant women in the Nile Delta, 37 were from chronic hepatitis patients and 38 were from cancer epidemiology studies from the National Cancer Institute located in Cairo, which represents the referral centre for all Egypt [18 pathologically verified HCC cases, 10 clinical HCC cases and 10 non-Hodgkin’s lymphoma (NHL) cases]. All subjects gave informed consent and were interviewed to obtain data on their socio-demographic characteristics and medical history. Institutional Review Board approval was obtained from all collaborating institutions in Egypt and the USA.

The method for sequencing of the mid-core region was based on a semi-nested approach described previously (Widell et al., 1994), modified to use inner primers that were 5'-tagged with 20 nt sequences to improve sequencing by using M13 and M13 reverse primers (for details, see Supplementary Table S1, available in JGV Online). For NS5B sequencing, we used a novel set of nested NS5B primers: JA230, 5'-CTACCATCATGGCTAA(A/G)AA(C/T)-GAGGT (outer sense, 8008–8032); JA233, 5’-ATGATGTATGAGCTCCA(A/G)GTCAG/AA(C/T)-GAGGT (outer antisense, 8663–8687); JA231, 5’-TATGA(C/T)ACCGCCTG(C/T)TCTTGAC (inner sense, 8256–8276); and JA232, 5’-CCTGGTCTAGCCCTGGAAA (inner antisense, 8616–8636). PCR details are presented in the Supplementary Material, available in JGV Online.

The mid-core and NS5B sequences were used for phylogenetic inference using the MEGA v. 3.1 software (Kumar et al., 2004) with the neighbour-joining method, Tamura three-parameter substitution model, gamma-distributed rates among sites and 1000 bootstrap replicates. MODELTEST v. 3.6 (Posada & Crandall, 1998) was used to obtain optimal gamma parameters (0.60 for mid-core and 1.2 for NS5B). Several other models were also used and gave similar topography. Reference sequences for HCV genotypes and subtypes were obtained from the Los Alamos HCV database (http://hcv.lanl.gov/content/hcv-db/index). For differences in observed versus expected frequencies across all clinical groups of subjects, we used $\chi^2$ tests (STATISTICA 5.1; StatSoft). Contingency tables were also assessed by $\chi^2$. P values of $<0.05$ were considered to be statistically significant.

The RFLP results showed that 95 samples had a typical HCV subtype 4a cleavage pattern (Mval/Hinfl) fragment sizes of 53, 56 and 128/129 bp and Rsal/HaeIII fragment sizes of 45, 58, 114, 9 and 11/12 bp or 103, 114, 9 and 11/12 bp). Fourteen samples displayed the same genotype 4 RFLP pattern with Mval/Hinfl, but were nontypable with Rsal/HaeIII, with fragment sizes of 217, 9 and 11/12 bp (absence of two Rsal cleavage sites) or 45, 172, 9 and 11/12 bp (absence of one Rsal cleavage site). We refer to these patterns as 4Unc (genotype 4 unclassified). Sixteen samples did not conform to the published Mval/Hinfl RFLP patterns and are referred to as Uncl (unclassified), giving fragment sizes of 53, 56, 49 and 79/80 bp (extra cut sites by Mval). Three samples displayed a typical genotype 1 RFLP pattern, one sample was genotype 3 and two samples suggested a mixture of two genotype 4 patterns, which we designated pattern 4Mix (mixed).

The phylogenetic tree based on the mid-core 217 nt (Fig. 1) showed that most samples clustered with genotype 4, whereas a few clustered with other genotypes: three in subtype 1b, five in genotype 1g (Ndjomou et al., 2003) and one in subtype 3a. Bootstrap support for genotype 4 was $<50\%$, in contrast to all other genotypes. Bootstrap support for most subtypes within genotype 4 was also low, with subtype 4a showing 73% and subtype 4o 63%. In general, the RFLP and mid-core genotyping results agreed well, but some exceptions were also noted. Thus, most samples with a subtype 4a RFLP pattern were located in a large genotype 4a cluster, which included well-known reference isolates such as ED43 (GenBank accession no. Y11604; Chamberlain et al., 1997). However, nine samples with the same RFLP pattern clustered with genotype 4n in the mid-core tree and one sample clustered within subtype 4o. Nine other samples in the subtype 4o cluster instead displayed the 4Unc RFLP pattern. Furthermore, clusters 4m and 4r were dominated by samples with the Unc1 RFLP pattern. These subtypes may also be mistyped by the commonly used reverse-hybridization assay as mixed type 1 and type 4 (Rapicetta et al., 1998). In addition, the large mid-core subtype 4a cluster also contained four samples with atypical RFLP patterns – two 4Uncl (LYM_EG_029 and LYM_EG_034) and two 4Mix (HCC_EG_014 and EG_055) samples. The five samples with subtype 1g (Ndjomou et al., 2003) showed the Uncl RFLP pattern and thus resembled samples that clustered with subtypes 4m, 4r and 4o according to the mid-core trees. Taken together, there was more diversity in mid-core sequences than described previously.
Amplification of NS5B was successful in 110 of the 124 samples for which a sufficient amount of serum was available. Thus, 21 samples were not included in NS5B analysis. Of the 14 specimens for which the NS5B analysis failed, seven were of subtype 4a, five 4n, one 4r and one 1b when sequenced in the mid-core region. Thus, there was a relative, but not absolute, weakness to amplify subtype 4n in this panel, possibly due to partial primer mismatch or, more likely, low RNA levels. For NS5B, a much larger array of well-defined typed reference strains was available, ranging from subtype 4a to 4r [terminology as suggested by the expert team at the Los Alamos HCV database and an international consensus group headed by Simmonds et al. (2005), where the recently described 4'beta' from Egypt (Ray et al., 2000) was renamed subtype 4o].

As shown in Fig. 2, the resolution of the NS5B tree was greater than that of the mid-core tree, as displayed by greater

---

**Fig. 1.** Neighbour-joining tree based on HCV mid-core sequences (217 nt) for 131 Egyptian (EG) and reference HCV sequences. Distances were calculated by the Tamura three-parameter algorithm using a gamma distribution of 0.60 obtained by using MODELTEST v. 3.6. Bootstrap values (of 1000 resamplings) above 50% are shown. The 131 isolates identified in this study are indicated by the code EG followed by a number. The prefix HCC denotes pathologically confirmed hepatocellular carcinoma, HCp denotes probable, but not pathologically confirmed HCC, and LYM denotes non-Hodgkin's lymphoma. The RFLP pattern is shown as follows: filled grey circle, RFLP pattern 4a; filled black circle, type 1; filled black square, type 4Uncl; filled black triangle, type Uncl; open square, type 4Mix (4+Uncl); open circle, type 3. Arrows indicate positions of isolates EG _ 011 and EG _ 023. Some subtype designations, such as subtype 4n, were inferred from the NS5B analysis because reference mid-core sequences for subtype 4n are lacking.
genetic distances and higher bootstrap support for established genotypes and subtypes. This is due to the greater variability and length of the NS5B fragment. Altogether, 108 samples gave concordant results between mid-core and NS5B, whilst two samples showed some discordance (EG_011 and EG_023). Sample EG_011 clustered closest to subtype 4l in the mid-core trees, but closest to subtype 4o in the NS5B trees, where it shared a bootstrap-supported branch with 4o isolates, despite only 85–87% identity with the 4o main cluster, which itself had 93–94% identity in NS5B and 98% bootstrap support. Sample EG_023 clustered closest to an isolate from Cameroon (GenBank accession no. AY256810; Ndjomou et al., 2003) in the mid-core tree, but in NS5B showed identity no greater than 86% to any sequence in GenBank, and may therefore represent a new subtype within genotype 4.

Our study did not reveal any recombination between subtypes or major genotypes, such as that published elsewhere (Kalinina et al., 2002). This argues against sample confusion or PCR contamination. However, it is not possible to exclude intrasubtypic recombination.

Sequencing of mid-core and NS5B confirmed the existence of subtype 1g sequences reported by Ray et al. (2000). The 1g sequences shared pattern UcInl with several isolates in subtypes 4m/4r and 4o.

Ray et al. (2000) focused on viral diversity in different governates in Egypt, using RFLP and sequencing in the core–E1 junction and, to a lesser extent, NS5B. Their findings on genetic distance are in line with ours, but we have added mid-core data, extended the NS5B database for Egypt and further added the malignancy aspect.

Our study and several others (Ray et al., 2000; Morice et al., 2001; Ndjomou et al., 2003) show that HCV genotyping is a challenge in highly endemic regions, such as Egypt. This has led to conflicting nomenclatures, as reviewed and resolved by Simmonds et al. (2005). We confirmed that more simple methods, such as RFLP, have limitations. In contrast, we found very good concordance between mid-core genotyping and the more widely used NS5B genotyping. However, it was obvious that sequence clusters were separated by greater genetic distances and supported by higher bootstrap values in NS5B trees than in mid-core trees. The two methods are very similar in terms of workload, expense etc. and we therefore feel that NS5B genotyping should be preferred over mid-core genotyping. However, a few samples that were genotyped successfully in mid-core could not be amplified with our novel NS5B primers.

5’-UTR sequences were obtained from selected samples to study further the discordant typing results obtained by the mid-core, NS5B and RFLP methods. In the 4o samples, the 5’-UTR sequences shared with 4m and 4n an adenosine insertion following position 206 (numbering according to the HCV-1 prototype; GenBank accession no. AF009606).

In genotype 4o, the pattern gaACAcAml (cluster-typical bases in bold upper case, insert underlined, common HCV -5’-UTR insert). We found very good concordance between mid-core genotyping and the more widely used NS5B genotyping.

### Table 1. Number (%) of verified (HCC) or clinically suspected (HCp) hepatocellular carcinoma and non-Hodgkin’s lymphoma (NHL) among the 131 samples analysed in this study

<table>
<thead>
<tr>
<th></th>
<th>HCC</th>
<th>HCp</th>
<th>NHL</th>
<th>All other subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>All genotype 1 (n=8)</td>
<td>3 (38)</td>
<td>1 (12)</td>
<td>0</td>
<td>4 (50)</td>
</tr>
<tr>
<td>Subtype 1b (n=3)</td>
<td>1 (33)</td>
<td>1 (33)</td>
<td>0</td>
<td>1 (33)</td>
</tr>
<tr>
<td>Subtype 1g (n=5)</td>
<td>2 (40)</td>
<td>0</td>
<td>0</td>
<td>3 (60)</td>
</tr>
<tr>
<td>All genotype 3 (n=1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 (100)</td>
</tr>
<tr>
<td>All genotype 4 (n=122)</td>
<td>15 (12)</td>
<td>9 (7)</td>
<td>10 (8)</td>
<td>88 (72)</td>
</tr>
<tr>
<td>Subtype 4a (n=83)</td>
<td>6 (7)</td>
<td>7 (8)</td>
<td>6 (7)</td>
<td>64 (77)</td>
</tr>
<tr>
<td>Subtype 4l (n=1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Subtype 4l (n=5)</td>
<td>2 (40)</td>
<td>0</td>
<td>0</td>
<td>3 (60)</td>
</tr>
<tr>
<td>Subtype 4m/r (n=10)</td>
<td>1 (10)</td>
<td>0</td>
<td>1 (10)</td>
<td>8 (80)</td>
</tr>
<tr>
<td>Subtype 4o (n=14)</td>
<td>6 (43)</td>
<td>2 (14)</td>
<td>3 (21)</td>
<td>3 (21)</td>
</tr>
<tr>
<td>Subtype 4n (n=9)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9 (100)</td>
</tr>
<tr>
<td>P value by $\chi^2$ for all groups in entire dataset</td>
<td>0.017</td>
<td>0.029</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>P value by $\chi^2$ for 4o versus all other patterns</td>
<td>0.0033</td>
<td>0.019</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

5’-UTR sequences were obtained from selected samples to study further the discordant typing results obtained by the mid-core, NS5B and RFLP methods. In the 4o samples, the 5’-UTR sequences shared with 4m and 4n an adenosine insertion following position 206 (numbering according to the HCV-1 prototype; GenBank accession no. AF009606).

In genotype 4o, the pattern gaACAcAml (cluster-typical bases in bold upper case, insert underlined, common HCV -5’-UTR insert). The other uncommon 5’-UTR, with pattern gaUuAaAml at the analogous positions around 206, was found in subtypes 4m (EG_026, EG_027 and HCC_EG_148) and 4n (EG_030 and EG_037), again differing from type 4a, which did not carry this 5’-UTR insert.

In our set of samples, we observed a significantly skewed HCV subtype distribution among the histologically confirmed HCC cases ($P=0.017$) (Table 1). This skew was slightly weaker if the 10 cases with probable HCC were added ($P=0.029$). Detailed analysis revealed that this
finding was mainly due to an overrepresentation of subtype 4o among HCC cases. Thus, six of 14 subtype 4o-infected patients had verified HCC and two additional patients had clinical HCC. HCC, as well as HCC and/or probable HCC, was observed significantly more frequently among patients infected with subtype 4o than in those infected with other subtypes ($P=0.0033$ and 0.019, respectively). Thus, in this panel, a significant association was observed between subtype 4o and HCC. In contrast, NHL was not associated significantly with any HCV subtype. As mentioned, the HCC cases were recruited from a reference clinic in Cairo that receives patients from all over Egypt, where subtype 4'beta'/4o is known to circulate in geographically distant regions (Ray et al., 2000). Previous studies on the possible association between genotype 1b and HCC have been contradictory (Nousbaum et al., 1995; Benvegnù et al., 1997) and it has been suggested that 1b cases may have been infected for longer. In our study, HCC cases also occurred in genotype 1-infected patients, but too few to draw further conclusions. Not surprisingly, HCC patients were older on average (54.3 years) than the mothers studied for vertical transmission (30.4 years) and the other community-based HCV carriers (39.6 years). However, among patients with verified HCC, the mean age in cases with 4o versus those with non-4o subtype was 55.9 versus 58.6 years, respectively. Thus, higher age alone can not explain the association between HCC and subtype 4o.

Although HCC cases were distributed among most genotype 4 and 1 subtypes, a significant and novel association
was observed between HCC and cluster subtype 4o. It is presently unknown whether subtype 4o has spread by other mechanisms or at earlier time points than the dominant subtype, 4a. As we originally did not design this study as a case–control study, the age of the patients was not matched between groups of subjects. This issue is now being addressed prospectively in a new case–control study that we are conducting in Egypt, where the cases and controls are being matched closely for age, sex, probable route and duration of HCV infection and geographical location. Hopefully, this study will provide a definitive answer on whether subtype 4o of HCV is associated significantly with an increased risk of HCC.

Acknowledgements

This work was supported by grants from the Swedish Cancer Foundation (grant 06-08-050510), ALF grants from the Medical Faculty of Lund University, the Cancer Foundation of Malmö University Hospital and the Alfred Osterlund Foundation. Additional support came from the US National Institutes of Health (grant RO1CA85888).

References


Morice, Y., Roulot, D., Grando, V., Strinemann, J., Gault, E., Jeantils, V., Bentata, M., Jarrousse, B., Lorholay, O. & other authors (2001). Phylogenetic analyses confirm the high prevalence of hepatitis C virus (HCV) type 4 in the Seine-Saint-Denis district (France) and indicate seven different HCV-4 subtypes linked to two different epidemiological patterns. *J Gen Virol* 82, 1001–1012.


