Survey of major genotypes and subtypes of hepatitis C virus using RFLP of sequences amplified from the 5' non-coding region

F. Davidson, 1 P. Simmonds, 2 J. C. Ferguson, 1 L. M. Jarvis, 2 B. C. Dow, 3 E. A. C. Follett, 3
C. R. G. Seed, 4 T. Krusius, 5 C. Lin, 6 G. A. Medgyesi, 7 H. Kiyokawa, 8 G. Olim, 9 G. Duraisamy, 10
T. Cuypers, 11 A. A. Saeed, 12 D. Teo, 13 J. Conradie, 14 M. C. Kew, 15 M. Lin, 16 C. Nuchaprayoon, 17
O. K. Ndimbie 18 and P. L. Yap 19

1 Edinburgh and South East Scotland Blood Transfusion Service, Edinburgh Royal Infirmary, Lauriston Place, Edinburgh EH3 9HB, UK, 2 Department of Medical Microbiology, University of Edinburgh, Edinburgh EH8 9AG, UK, 3 Scottish National Blood Transfusion Service Microbiology Reference Laboratory, Ruchill Hospital, Glasgow, UK, 4 Australian Red Cross Society, Blood Transfusion Service, Perth, Australia, 5 Finnish Red Cross Blood Transfusion Service, Helsinki, Finland, 6 Hong Kong Red Cross Red Cross Blood Transfusion Service, Yaumatei, Hong Kong, 7 National Institute of Haematology, Blood Transfusion and Immunology, 1113 Budapest, Hungary, 8 Fukuoka Red Cross Blood Transfusion Centre, Fukuoka 818, Japan, 9 Macau Blood Transfusion Centre, Av. Sidonio Pais 49 R/C AS, Macau, 10 National Blood Services Centre, Hospital Kuala Lumpur, Jalan Pahang, 50586 Kuala Lumpur, Malaysia, 11 Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, 1066 CX Amsterdam, The Netherlands, 12 Riyadh Armed Forces Hospital, Riyadh 11159, Saudi Arabia, 13 Singapore Blood Transfusion Service, National Blood Centre, Outram Road, Singapore 0316, 14 The Natal Institute of Immunology, Division of The Natal Blood Transfusion Service, PO Box 2356, Durban 4000, South Africa, 15 University of Witwatersrand, Medical School, Parktown, Johannesburg, South Africa, 16 Transfusion Medicine Laboratory, National Health Research Institute, Mackay Memorial Hospital Tapei, Taiwan, 17 National Blood Centre, Thai Red Cross Society, Bangkok 10330, Thailand, 18 Central Blood Bank, 812 Fifth Avenue, Pittsburgh, Pa., USA

A method is described for identifying different genotypes of hepatitis C virus (HCV) by restriction endonuclease cleavage of sequences amplified by PCR from the 5' non-coding region. Using the enzymes HaeIII–RsaI and HinfI–MvaI, followed by cleavage with BstU1 or SfeI, it was possible to identify and distinguish HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5 and 6. The method was used to investigate the prevalence of these genotypes in 723 blood donors in 15 countries, the largest survey to date, and one which covered a wide range of geographical regions (Europe, America, Africa and Asia). These results, combined with a review of the existing literature, indicate the existence of several distinct regional patterns of HCV genotype distribution, and provide the framework for future detailed epidemiological investigations of HCV transmission.

Introduction

In order to understand the epidemiology of hepatitis C virus (HCV) infection, the identification of virus genotypes provides a useful method to compare HCV in different geographical regions and in different age, risk and racial groups (McOmish et al., 1994; Dusheiko et al., 1994; Murphy et al., 1994; Woodfield et al., 1994; Kleter et al., 1994; Lee et al., 1994; Tismanetzky et al., 1994; Viazov et al., 1994). According to recent proposals, variants of HCV can be classified into a total of six major genotypes, many of which contain a number of more closely related subtypes (Simmonds et al., 1993a, 1994a; Homes et al., 1993; Bukh et al., 1993). Initial findings indicate major biological differences between genotypes: for example, preliminary studies show that individuals infected with different genotypes respond differently to treatment with interferon (reviewed by Dusheiko & Simmonds, 1994). Should these findings be confirmed, the existence of geographical variation in the predominant circulating genotypes in different communities would lead to the adoption of different treatment policies for patient management. Furthermore, the documented antigenic differences between HCV genotypes (McOmish et al., 1993, 1994) would have implications for blood donor screening and for the design and composition of future candidate vaccines for HCV in different countries.

We have previously described a method by which the different major genotypes of HCV may be identified by...
cleavage of the amplified 5' non-coding region (5'NCR) using restriction enzymes that recognize sequence polymorphisms between HCV genotypes (McOmish et al., 1993). This method suffered from two major drawbacks in comparison with other typing methods. Firstly, although it could reliably identify all six major genotypes of HCV, it was unable to separately identify HCV subtypes, which themselves might show important epidemiological or biological differences. Secondly, the method was dependent upon the electrophoresis of radiolabelled DNA fragments on acrylamide gels involving considerable expenditure on equipment and radioisotopes, and being technically relatively difficult and time-consuming.

In the current study, we have modified the assay to allow the identification of subtypes 1a, 1b, 2a, 2b, 3a and 3b, as well as types 4, 5 and 6. DNA fragments were separated on high resolution agarose gels, so that the need for radioisotopic labelling of DNA was eliminated. This modified assay was used to identify the main genotypes infecting populations in a total of 15 countries, representing Europe, Asia, Australia, Africa and the USA. The method described is a rapid and reliable method for surveying HCV genotype distributions in a wide range of settings.

Methods

Samples. Samples used for this survey were obtained from 723 volunteer blood donors from 15 countries. Donors were in each case screened using second generation enzyme immunoassays (EIAs); samples that were reactive were amplified by PCR as described below. Details of samples used from Scotland, Netherlands, Finland, Hungary, Australia, Japan, Taiwan, Hong Kong and Egypt have been previously described (McOmish et al., 1994). The additional samples from the USA, Macau, Singapore, Thailand and Malaysia were screened by Abbott second generation EIA. Donors in Johannesburg, South Africa, were screened by UBI ELISA, while donors in Durban, South Africa, were screened by a combination of Ortho second generation and UBI ELISA as previously described (Kew et al., 1990).

Samples from the USA, Singapore and Thailand were not subjected to confirmatory antibody testing; for the USA, 119 repeat reactive samples yielded 54 that were PCR-positive; for Singapore 11/40, and for Thailand 77/90 tested were PCR-positive; however, one sample from Thailand and one other from Singapore failed to reamplify from
storage and are therefore not included in the table of genotyping results. Confirmed or indeterminate samples upon confirmatory testing were used from Malaysia (Abbott Matrix: 25/39 PCR-positive), Macau (RIBA-2: 27/57 PCR-positive) and from South Africa (RIBA-2: Durban: 25/29; Johannesburg: 63/84).

RFLP analysis. RNA was extracted from 100 μl of serum, reverse transcribed and amplified by nested PCR using 5'NCR specific primers 939, 209, 940 and 211 as previously described (Chan et al., 1992). Restriction digests were carried out on 25 μl of (unlabelled) secondary PCR product for 4–16 h after adjustment with 10× enzyme reaction buffer as appropriate. Reactions were at 37 °C in the presence of 10 units each of (a) RsaI and HaeIII, (b) HinfI and MvaI, (c) ScrFI or (d) BstUI (at 60 °C). Digestion products were visualized under UV light after electrophoresis through a 4% Metaphor agarose gel (FMC Bioproducts) in 1× Tris-borate buffer (134 mM-Tris-HCl, pH 10.0; 68 mM-boric acid; 2.5 mM-EDTA) containing 0.5 μg/ml ethidium bromide (Fig. 1).

Results and Discussion

Predicted electropherotypes of published HCV 5'NCR sequences of known genotype

Previous work has demonstrated that all six currently defined major genotypes of HCV (Simmonds et al., 1993a; Bukh et al., 1993) can be distinguished on the basis of RFLP analysis of RT-PCR products derived from the 5'NCR (Nakao et al., 1991; McOmish et al., 1993, 1994; Simmonds et al., 1994b). In one previous study, amplified DNA is cleaved with HaeIII–RsaI in one reaction and with MvaI–HinfI in a second, and the virus genotype is deduced by the combination of electropherotypes produced (McOmish et al., 1994). Comparison of published 5'NCR sequences also revealed the existence of sequence polymorphisms that are associated with individual HCV subtypes (Stuyver et al., 1993; Simmonds et al., 1994b), raising the possibility of extending the RFLP assay in order to discriminate between subtypes.

For subtypes 1a and 1b (following the nomenclature proposed by Simmonds et al. (1994a)) an association has been described between subtypes 1a and 1b and a sequence polymorphism at position −99 (numbering as in Choo et al., 1991) with an adenosine (A) residue in type 1a viruses and a guanosine (G) in subtype 1b. This association is not absolute since in a survey of 34 type 1 5'NCR sequences the G polymorphism is present in one type 1a virus (DR4) (Simmonds et al., 1994b). The presence of a G at position −99 produces a sequence in RT–PCR products that is recognized by the restriction enzymes BstUI and MvnI. There are no other polymorphisms that affect this restriction site amongst the type 1 viruses surveyed, except for a single type 1a virus (DK9) that has a C to U substitution at position −100 (Bukh et al., 1992, 1993) and in any case lacks the G at position −99. Accordingly, the subtype of type 1 viruses can usefully be predicted on the basis of the RFLP pattern produced after digestion with BstUI resulting in DNA fragments of 209 and 42 bp for type 1a viruses and 179, 42 and 30 bp for type 1b viruses (Fig. 2f). Although the results are presented as either subtype 1a or 1b, as these are the principle subtypes found in most populations, it should be borne in mind that a small proportion of variants may correspond to other subtypes of type 1. For example, type 1c (Simmonds et al., 1993b) would appear as type 1a in this assay.

Subtype-specific polymorphisms have also been described for the 5'NCR of type 2 viruses (Stuyver et al., 1993), with the consistent presence of a C to U substitution at position −124 in 9 type 2b viruses, as well as a G to A substitution at position −161 in 8 of these sequences (Simmonds et al., 1994b). Both of these changes disrupt ScrFI restriction sites that are present in all eight type 2a 5'NCR RT–PCR products, so that type 2b viruses are expected to show ScrFI–HinfI electropherotypes E (eight viruses) or F (one virus), with all eight type 2a viruses having electropherotype D (McOmish et al., 1994). Since there are no HinfI restriction sites in the type 2 5'NCR sequence, the same electropherotypes will be produced by digestion with ScrFI alone (Fig. 2i). Type 2c, a third subtype of type 2 (Simmonds et al., 1993a; Bukh et al., 1993), has a 5'NCR sequence similar to that of type 2a viruses and three out of four sequences are expected to give ScrFI electropherotype D and one electropherotype E (Simmonds et al., 1994b). Thus, types 2a and 2b can be distinguished from each other by RFLP analysis with ScrFI, while type 2c viruses may be misclassified as either type 2a or 2b. Additional sequence data will help to establish whether it would be possible in the future to modify the assay to allow type 2c to be separately identified.

Few 5'NCR sequences are available for type 3 subtypes other than type 3a (Tokita et al., 1994). All 16 type 3a 5'NCR sequences have a C to U substitution at position −164 and a G to A substitution at position −121 (Simmonds et al., 1994b; Tokita et al., 1994) destroying ScrFI restriction sites. These substitutions are also present in single examples of types 3c, 3d and 3e but are not found in the two sequences from types 3b (Tokita et al., 1994) (K. Chayama and others: GenBank accession number D114443), and in 3f (Tokita et al., 1994), producing ScrFI electropherotypes G and A respectively (McOmish et al., 1993) (Fig. 2iii).

Survey of HCV genotypes in blood donors

RNA from samples of 723 blood donors from 15 countries was amplified in the 5'NCR, and the product cleaved with HaeIII–RsaI and MvaI–HinfI in two separate reactions. Those samples identified as type 1
were further examined by cleavage with BstUI, while type 2 and type 3 samples were cleaved with ScrFI. The cleavage patterns of the DNA fragments were determined by agarose gel electrophoresis; examples of some of the restriction patterns are shown in Fig. 1.

**Europe.** Types 1, 2 and 3 were the predominant genotypes infecting blood donors throughout Europe. In particular, there was a relatively high frequency of type 1a and 3a infection in the north-western countries (Scotland, Eire and Finland) that contrasts with the more frequent type 1b infections in blood donors and patients from Hungary (Table 1), Germany, Russia and Turkey (Viazov et al., 1994), and in blood transfusion recipients in Estonia (unpublished data).

We have previously remarked upon the similarity in genotype distribution between Australia (Perth) and Scotland or other northern-European countries (McOmish et al., 1994). Using the modified typing assay, we have now found that the relative frequencies of type 1a to 1b, and of 2a to 2b were also similar. In North America, type 1a was the most frequent genotype infecting blood donors in Pittsburgh; varying proportions of type 1a to 1b have been found in NANB patients elsewhere in the USA (Lau et al., 1994) and in Canada (Andonov & Chaudhary, 1994) [65% and 34% (this study); 68% and 82% in other studies respectively], while type 3a infection was rare (4.7%), consistent with previous reports (Lau et al., 1994; Andonov & Chaudhary, 1994; Murphy et al., 1994). Although types 1a and 1b have been described as the ‘American’ and ‘Japanese’ types (Pozzato et al., 1991, 1994), these terms are misleading as type 1b accounts for 13–27% of HCV infections in North America, and is the predominant genotype in many countries in Europe (McOmish et al., 1994; Viazov et al., 1994). Similarly, type 1a is not confined to America.

**Far-East and South-East Asia.** Major intra-regional differences in genotype distribution are apparent. In Japan and Taiwan, types 1b, 2a and 2b were the only variants detected (Table 1, Fig. 3). The absence of 1a contrasts strongly with Europe and North America, but is consistent with previously published investigations, where it has been reported that this genotype is only found in haemophiliacs treated with imported (USA origin) factor VIII and IX concentrates (Hijikata et al., 1990; Kinoshita et al., 1993). We previously reported the finding of a high frequency of type 6a infection in Hong Kong (McOmish et al., 1994). We have now found this genotype to also account for approximately 30% of HCV infections in blood donors in Macau; there is now preliminary evidence that this genotype is more widely distributed in this region, for example in Vietnam.

---

### Fig. 2. Predicted 5'NCR electropherotypes for restriction enzymes used for HCV typing.

(i) BstUI cleavage of type 1a and 1b sequences producing patterns A and B respectively; (ii) ScrFI of types 2a and 2b, producing patterns D (type 2a), E and F (type 2b) (McOmish et al., 1993); (iii) ScrFI cleavage of type 3a and 3b sequences, producing cleavage patterns G and A respectively. Examples of cleavage patterns produced upon agarose gel electrophoresis of fragments are shown in Fig. 1.
Table 1. Genotype distributions in 15 countries identified by RFLP assay

<table>
<thead>
<tr>
<th>Region</th>
<th>Country</th>
<th>Centre</th>
<th>No. of samples</th>
<th>HCV genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1a* 1b 2a† 2b 3a‡ 3b 4 5 6</td>
</tr>
<tr>
<td>Europe</td>
<td>Scotland</td>
<td>(National)</td>
<td>144</td>
<td>56 11 16 5 56 - - - -</td>
</tr>
<tr>
<td></td>
<td>Finland</td>
<td>(National)</td>
<td>12</td>
<td>2 1 - 5 4 - - - -</td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td>Amsterdam</td>
<td>31</td>
<td>6 12 5 2 5 1 - - - -</td>
</tr>
<tr>
<td></td>
<td>Hungary</td>
<td>(National)</td>
<td>45</td>
<td>6 38 - - - - - - - -</td>
</tr>
<tr>
<td>Far-East</td>
<td>Taiwan</td>
<td>Taichung</td>
<td>93</td>
<td>1 52 28 12 - - - - - -</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>Fukuoka</td>
<td>40</td>
<td>- 31 8 1 - - - - - -</td>
</tr>
<tr>
<td></td>
<td>Hong Kong</td>
<td>NA</td>
<td>35</td>
<td>1 21 - 1 - - - - - 12</td>
</tr>
<tr>
<td></td>
<td>Macau</td>
<td>NA</td>
<td>27</td>
<td>2 13 3 1 4 - - - - 6</td>
</tr>
<tr>
<td>South-East Asia</td>
<td>Singapore</td>
<td>NA</td>
<td>10</td>
<td>1 4 2 - 3 - - - - -</td>
</tr>
<tr>
<td></td>
<td>Malaysia</td>
<td>Kuala Lumpur</td>
<td>25</td>
<td>8 8 - - - 9 - - - -</td>
</tr>
<tr>
<td></td>
<td>Thailand</td>
<td>Bangkok</td>
<td>76</td>
<td>10 24 - - - 36 6 - - -</td>
</tr>
<tr>
<td>Other</td>
<td>Australia</td>
<td>Perth</td>
<td>24</td>
<td>9 3 2 1 8 - - - 1 -</td>
</tr>
<tr>
<td></td>
<td>South Africa</td>
<td>Johannesburg</td>
<td>63</td>
<td>3 16 1 - - 2 - 4 37 -</td>
</tr>
<tr>
<td></td>
<td>Durban</td>
<td>25</td>
<td>2 5 1 - - 5 - - - - 12 -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Pittsburg</td>
<td>54</td>
<td>27 14 3 6 3 - - - 1 -</td>
</tr>
<tr>
<td></td>
<td>Egypt</td>
<td>NA</td>
<td>19</td>
<td>- - - - - - - - - 19 -</td>
</tr>
<tr>
<td></td>
<td>Total...</td>
<td></td>
<td>723</td>
<td>134 251 69 34 135 7 24 51 18</td>
</tr>
</tbody>
</table>

* Variants producing cleavage pattern A with BstUI are referred to as type 1a, while those producing pattern B are referred to as type 1b. Other subtypes of type 1 produce pattern A (see text).
† Variants producing cleavage pattern D or E/F with ScrFI are referred to as type 2a or 2b. The four known type 2c sequences produce patterns D or E (see text).
‡ Variants producing cleavage pattern G or A with ScrFI are referred to as type 3a or 3b (see text).
NA, Not applicable.

Fig. 3. Cartographic representation of distribution of HCV genotypes in the Far-East. Pointers indicate the location of the Blood Donor Centre in the countries participating in the survey. Graphs show numbers (y-axis) of each of the six major genotypes of HCV (x-axis). For types 1, 2 and 3, numbers of subtypes are shown by shading of the histogram: [●], subtype a; [●], subtype b).
This paper describes the optimization and simplification of an existing typing method based upon RFLP of the 5'NCR for identification of genotypes. In common with other techniques, such as the use of type-specific probes in the 5'NCR (Stuyver et al., 1993), or the use of genotype-specific PCR primers in the core region (Okamoto et al., 1992), the RFLP method is dependent upon the existence of a relatively small number of nucleotide sequence differences between known examples of different genotypes. One of the advantages of using the 5'NCR for genotyping is the existence of a few identifiable polymorphisms between major genotypes that reliably associate with examples of the known genotypes (Simmonds et al., 1994b), and for which it is relatively straightforward to design combinations of restriction enzymes that recognize them. Furthermore, the 5'NCR is the region of choice for detection and quantification of HCV in vivo, and the RFLP assay can be conveniently added on to conventional PCR methods used for diagnosis.

It is likely that typing methods (including RFLP) will have to be continuously modified to accommodate the rapidly increasing database of information collected on HCV sequence heterogeneity. The existence of numerous subtypes of type 1 (Okamoto et al., 1994; Simmonds et al., 1993b), 3 (Tokita et al., 1994) and 4 (Bukh et al., 1994), and the description of new major genotypes in South-East Asia (Hotta et al., 1994; Apichartpiyakul et al., 1994; Hadiwandowo et al., 1994) hints at a complexity in genotype distribution not generally found in Western countries, and which potentially affects the accuracy of any survey based upon typing assays. For example, recent reports describe the detection of HCV variants in Thailand and Indonesia that differ from those classified as genotypes 1 to 6 (Tokita et al., 1994; Hadiwandowo et al., 1994). We therefore have to be cautious in the interpretation of the results from Thailand, Macau, Singapore and possibly Taiwan reported here, as it is possible that a proportion of individuals may be infected with novel variants of HCV that produce RFLP patterns resembling other genotypes.

However, the existing assay provides an accurate and inexpensive method for the identification of HCV genotypes 1 to 6 in population screening, and is sufficiently accurate in most populations to be of value as part of pre-treatment assessment of patients undergoing interferon treatment. Our ongoing work is designed to further refine the assay to allow the identification of further HCV genotypes and subtypes, and to extend the survey to areas of the world where little information is currently available.

**Notes added in proof.** Since submission of the manuscript, novel HCV variants that produce RFLP patterns corresponding to type 1a or type 1b have been found in Thailand, similar to those reported from Vietnam. The extent and geographical distribution of these new genotypes in south-east Asia are currently being investigated using an RFLP method based upon the core region.


(Received 31 October 1994; Accepted 13 January 1994)