BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **236,** 44–49 (1997)
ARTICLE NO. RC976902

The Complete Coding Sequence of Hepatitis C Virus Genotype 5a, the Predominant Genotype in South Africa

R. W. Chamberlain,* N. J. Adams,† L. A. Taylor,* P. Simmonds,† and R. M. Elliott*
**Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, United Kingdom; and †Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom*

Received May 30, 1997

Hepatitis C virus (HCV) genotype 5a is the predominant genotype in southern Africa with a high prevalence amongst infected blood donors from areas in South Africa. We have determined the nucleotide sequence corresponding to the complete coding region of an HCV isolate, EUH1480, previously classified as genotype 5a, from an Edinburgh haemophiliac. The sequence contained a single open reading frame (ORF) coding for a polyprotein of 3014 amino acids. Comparison with the polyprotein sequences from other HCV genotypes, where the ORF varies from 3008 to 3037 amino acids, showed the observed variation in size was due to differences in lengths of the envelope 2 and the nonstructural 5A proteins. The sequence divergence of HCV genotype 5 ranged from 29.4% nucleotide differences (24.9% amino acid differences) compared with genotype 1c to 32.5% nucleotide differences (30.3% amino acid differences) compared with 2a. Phylogenetic analysis of the available full length nucleotide sequences showed EUH1480 to form a branch distinct from the other HCV types, confirming the classification of type 5a as a separate genotype.

Hepatitis C virus (HCV) is a member of the *Flaviviridae* family, which contains viruses characterised by having a single-stranded, positive sense RNA genome. A single open reading frame (ORF) encodes a polyprotein which is processed by host and viral encoded proteases. HCV isolates have been classified into six main groups (genotypes) which are subdivided into a varying number of subtypes (a, b, c *etc*) in each (1,2). This classification is based on phylogenetic analysis of partial sequences from the 5′ non-coding region (NCR) (3,4), the core (5,6), envelope 1 (6,7,8), and the nonstructural 5B genes (6,8,9), as well as from complete genomic sequences where available. In addition to the six main genotypes, new variants from Indonesia, Vietnam and Thailand have been described as genotypes 7, 8, 9, 10 and 11 (refs. 10-12). However, subsequent re-evaluation of these data has led many authors to suggest that types 7, 8, 9 and 11 would be more usefully described as subtypes of genotype 6, and genotype 10 as a variant of genotype 3, thus confining classification of HCV isolates to the originally proposed six genotypes (13-16).

Genotypes 1, 2 and 3 are the predominant genotypes found in Europe, USA and with the exception of genotype 3, the Far East and parts of Africa, whereas on the Indian subcontinent and Thailand genotype 3 alone is the predominant genotype. Genotype 4 is the principal type in the Middle East, north and central Africa. Genotype 5 shows a narrow range of distribution being restricted to southern Africa with a high prevalence (30% to 50%) amongst HCV-infected blood donors in South Africa (17). Relatively little sequence heterogeneity is found within this genotype and thus far only a single subtype has been described (5,8,9,17,18,19). Genotype 6 also demonstrates a narrow distribution and has only been found in SE Asia around Hong Kong, Macau and Vietnam (20,21).

Complete genomic sequences have been determined for isolates representing three subtypes (a, b and c) each from genotypes 1 and 2, two subtypes (a and b) from genotype 3, and the putative genotypes 10 and 11. Recently we reported the complete genomic sequences of isolates belonging to genotype 4 (ref. 22) and genotype 6 (ref. 16). Genotype 5 is the remaining genotype where the determination of the complete genomic sequence is outstanding. This paper addresses this by reporting the sequence of the complete ORF, up to the 3′ NCR, for an HCV isolate, EUH1480, identified as genotype 5a, from an Edinburgh haemophiliac. Partial sequence data have been reported for this genotype (5,8,9), but now we complete the genomic sequence for genotype 5 and therefore make available complete genomic sequences for representatives of all six major genotypes.

MATERIALS AND METHODS

Generation of PCR fragments. HCV RNA was extracted from 100 μl of plasma from an Edinburgh haemophiliac identified as genotype

0006-291X/97 $25.00

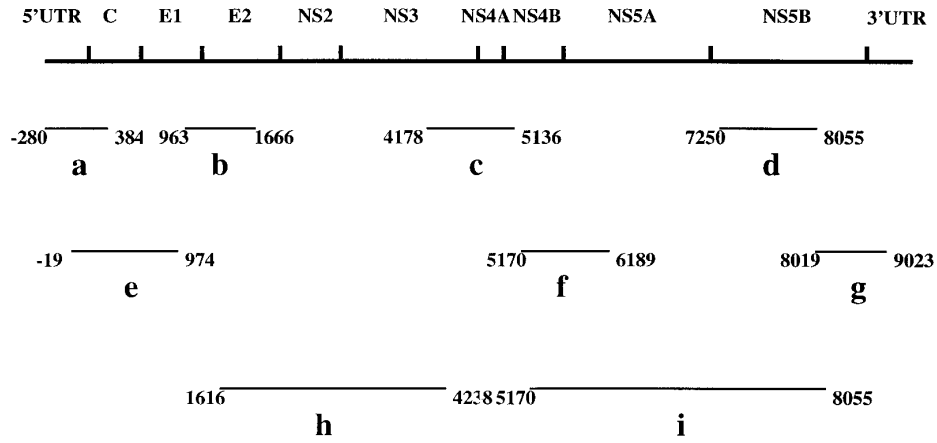

FIG. 1. Diagramatic representation of the HCV genome showing the position of the PCR fragments used to determine the complete sequence of the genotype 5. Nucleotide co-ordinates are based on the HCV-PT sequence (32).

5 and amplified by reverse transcription PCR using the method described previously (16). Primer sequences were either based on conserved regions as used previously or were genotype 5 specific based on type 5 sequence as it became known. All primer sequences are available on request. The amplification cycles for fragments *a-g* (Fig 1) were: 94°C for 18 secs, 45°C for 21 secs and 72°C for 90 secs using thin walled tubes on a Gene-E thermal cycler (Techne) with a hot start for the primary PCR (80°C for 2 mins). Fragment *h* was generated using a hot start (94°C for 1 min) followed by cycles of 94°C for 25 secs, 65°C for 30 secs and 68°C for 3 mins. To amplify fragment *i* the reverse transcription reaction products were incubated (37°C for 20 mins) with 3U of RNase H (Gibco BRL) and 1400U RNase T_1 (Gibco BRL) and then the PCR conditions were: 15 cycles of 96°C for 35 secs, 67°C for 30 secs and 68°C for 9 mins, followed by 10 cycles of 96°C for 35 secs, 67°C for 30 secs and 68°C for 11 mins and 15 cycles of 96°C for 35 secs, 67°C for 30 secs and 68°C for 10 mins with a final incubation of 72°C for 30 mins. Prior to amplification a hot start was carried out at 94°C for 1 min. Amplification reactions for these two fragments were performed in thin-walled tubes on a Progene thermal cycler (Techne) and based on the method described by Tellier *et al.* (ref. 23) using the Advantage Klentaq polymerase mix (Clontech).

CLONING AND SEQUENCING

All PCR fragments were electrophoresed in a 1% agarose TAE gel, visualised by ethidium bromide staining and UV illumination and the DNA purified by glass milk extraction (Geneclean; Bio 101). Fragments *a-h* were ligated overnight into plasmid pTAG and used to transform competent cells (pTAG Ligator kit; R&D Systems). Fragment *i* was end-repaired and blunt-end cloned into *Sma* I-digested pUC119. Three clones were picked for each PCR fragment and DNA prepared using the Qiagen plasmid mini prep kit with Qiagen tip 20's and sequenced using the Prism dye terminator cycle sequencing ready reaction kit (ABI) on a Perkin Elmer geneamp PCR system 2400 thermal cycler. The sequences were determined using a Prism 377 automated sequencer (ABI) and analysed using the genome assembly program (GAP) from the Staden package (24). Phylogenetic analysis was carried out using the DNADIST and NEIGHBOR programs from the PHYLIP package (25). The sequence of HCV-genotype 5 has been deposited in the Genbank database under the accession number Y13184.

RESULTS AND DISCUSSION

Three clones from each PCR fragment were sequenced on both strands and the consensus taken as representing the authentic sequence of EUH1480. A 9343 nucleotide continuous stretch of sequence was produced originating from the 5′ NCR and continuing to the start of the 3′ NCR (Fig. 1). Analysis of this sequence revealed a single open reading frame encoding a polyprotein of 3014 amino acids (aa). Comparison of the amino acid sequence with those of other HCV genotypes, which vary in length from 3008 aa for type 4a to 3037 aa for type 2c, showed the difference in size between genotypes to be attributable to differences in length of the E2 and NS5A proteins.

The nucleotide sequence from genotype 5 shows approximately 30% or more divergence from the other genotypes. Genotype 1 shows greatest similarity with a mean divergence of 30.1% with 1a (range of 29.9%-30.4%; n=4), 29.9% with 1b (range 29.7%-30.1%; n=13) and 29.4% divergence with 1c, whereas genotype 2 is the most distantly related with 32.5% divergence from genotype 2a, 32.3% from genotype 2b and 32.5% from genotype 2c. The type 3 sequences (3a and jk049) show 31.5% divergence and 31.7% (type 3b) whereas genotypes 4 and 6 show 30.9% and 31.6% divergence (30.7% for jk046) respectively.

The availability of the type 5 coding sequence allowed a more complete comparison of variability at a number of functional sites, such as the targets for the HCV NS2/NS3 and NS3 proteases (Fig 2A), the puta-

(A)

Genotype	NS2/NS3		NS3/NS4A		NS4A/NS4B		NS4B/NS5A		NS5A/NS5B	
1a HCV-PT	ADGMVSKGWRLL	APITAY	MTCMSADLEVVT	STWVLV	VLYREFDEMEEC	SQHLPY	HQWISSECTTPC	SGSWLR	SSEANAEDVVCC	SMSYSW
1a HCJ1	.....	.....	.....	.....	.....	.....	...L...S....	.....	....GT.....	....T.
1b HCV-J	..SFGEQ.....	.....	.A.....	.....	...Q.....	AS....	....NED.S...	.....K	.G..-G.....	....T.
1b BK	..SLEGR.L...	.....	.A.....	.....	L..Q.....	AS....	....NED.S...	.....	.E..-S.....	....T.
1c HCJ9	..ALTD.....	.....	.A.....	.....	.....	AA.I..	.V....D..A..	A....K	..DGGT.....	.....
2a HCJ6	...YT....S..	.....	A...Q.....M.	.....A	...EA.....	ASRAAL	.N..TED.PI..	.....	.E.--DDS....	.....
2b HCJ8	...YT....K..	.....	A...Q....IM.	.S...A	I..EA.....	ASKAAL	.A..TED.PV..	.....Q	.DQ--EDS.I..	.....
2c BEBE1	...YT.....	.....	A...Q....IM.	.....A	...EA.....	ASRTAL	....TED.PV..	.....	.DV--DDS....	.....
3a NZL1	..DYREM.....	.....	.A.....T.	L	...QQY.....	..AA..	....NEDYPS..	.DD...	.DSE-EQS....	.....
3a HPCHK6	..DYREM.....	.....	.A.....T.	L	...QQY.....	..AR..	....NEDYPS..	..D...	.DSE-EQS....	.....
3b Tr	..DYKKM.....	S..S..	.A.....T.	.A....	...QQY.....	..SA..	....NEDYPS..	N.D..H	.DSE-EQS....	.....
10a JK049	..DYR.M.....	.....H	A...A.....A.	.A...L	...QQY.....	.RAA..	....NEDYP...	D.N..Y	.G.--EQS....	.....
4a ED43	..TET.....	.....	.A.....	.....	...QQ.....	.K...L	.K..NED.S...	AE...W	.G---S.....	.....
5a EUH1480	..DIKTS.....	.....	.A.....I.	.....	...QQ.....	.AS...	.T..GEDYS...	D.T...	.G.---DN....	....T.
6a EUHK2	..D.QRG..K..	.....	.....I.	.....	I..QQ.....	.R.I..	...VNEDTA...	AT....	.D.---D.....	.....
11a JK046	..DVKD...S..	.....	.A.....I.	....VA	.M.QQY.....	.R....	.H..TEDYA...	GST...	...--ETS....	.....

1015 1032 1646 1663 1700 1717 1961 1978 2408 2425

(B)

Genotype	IFN sensitivity determining region
1a HCV-PT	PSLKATCTANHDSPDAELIEANLLWRQEMGGNITRVESEN
1a HC-J1	........I.....
1b HCV-J	........TH......D.....
1b BK	........TH.V....D.....
1c HCJ9	........TH......D..T.....
2a HC-J6	...R....THGKAY.VDMVD...F-----...DV..I...S
2b HC-J8	........THKTAY.CDMVD...F-----...DV..I..DS
2c HEBE1	...R....THAKC..IDMVD...FCWCT....M..I...S
3a NZL1	.......QTHRPH.....VD.....S.....T
3a HPCHK6	.......QTHRPH.....VD.....S.....T
3b Tr	.......QTHRPH.....D.....S.....T
10a JK049	..R.....THGRH.....T.....S.....S
4a ED43	RL.Q.....P....GTD.L.....----.STA....TDE
5a EUH1480	........IQGHH...D..K.....C.....A..
6a EUHK2	...R....TSQKH.EM..LQ.....KH...SH.P..Q...
11a JK046	........THA.H.....V.....

2209 2248

(C)

E1	E2 hypervariable region	E2
AKVLVVLLLFAGVDA	ETHVTGGSAGHTVSGFVSLLAPGAK	QNVQLINTNGSWHLN
.....	..I.S..Q.ARAM..L...FT....	..I.....I.
....I.M.....G	H.....RVASSTQSL..W.SQ.PS	.KI..V.....I.
....I.M.....G	D.....AQAK.TNRL..MF.S.PS	.KI.....I.
...V.....	..R....A....AF..A.F.....	.KI.....I.
...V.I...A.....	Q..TV...TA.NARTLTGMFSL..R	.KI.....I.
...IAI...V.....	T.YSS.QE..R..A..AG.FTT...	..LY.....I.
...V.I...T...E.	S.YT..AVV.RSTHL.T.MFSL.SQ	.R....H.....I.
...AIIMVM.S....	H.YT...T.SRHTQA.AG.FDI.PQ	.KL..V.....I.
...AIIMVM.S....	S....A.Q.ARNAY.IT..FSV...	..L.....I.
...VI..IM.S....	T..T.....AQATA..T.FFTR.PS	..L..V.S.....I.
...FL..C..S....	S.TI...V.ASGAFTIT..FST...	.PLH.V.....I.
...IL..F.....	....S.AAV.RSTA.LAN.FSS.S.	..L....S.....I.
...IL..F.....G	R..TV..TV.QGLKSLT.FFN..PQ	RQL.FV.....I.
L...A..F.....E.	Q.MI-AHGVSQ.T...A...T....	..I.....I.
...II.....	S.YV-AS.VSQAT..L...FSA..R	..L.....I.

369 384 408 423

FIG. 2. Comparison of inferred amino acid sequences (numbered as in Choo *et al.*, ref 32) of HCV genotypes 1–6. (A) Proteolytic cleavage sites mediated by NS2/3 and NS3 proteases shown from positions P12 to P6′ (cleavage occurs between P1 and P1′, as indicated by an arrow); P6, P1 and P1′ residues corresponding to the consensus sequence D/E -X_4- T/C ↓ S/A indicated in bold. (B) Putative interferon sensitivity determining region in NS5A (ref. 30). (C) Hypervariable region and flanking regions of E1 and E2; conserved amino acids residues are indicated in bold.

tive interferon sensitivity determining region in NS5A (Fig 2B) and the hypervariable region in E2 (Fig 2C). Cleavage between NS2 and NS3 is mediated by a virally-encoded protease comprising NS2 and the N-terminal one third of NS3. This enzyme has been shown experimentally to be tolerant of a wide range of non-conservative amino acid substitutions at positions P5-P3′ (26). The carboxyl terminal amino acids residues are identified by Pn and the amino terminal residues by Pn′, cleavage therefore, occurs between P1 and P1′. Nevertheless, the cleavage site for this enzyme is extremely well conserved between genotypes with each sequence corresponding closely to the consensus sequence GWRLL ↓ API (Fig 2A).

Considerable sequence variability was observed, however, between genotypes at sites critical for recognition by the HCV serine protease, NS3, although sequences were highly conserved within a genotype. The minimal recognition sequence for NS3 was proposed to be D/E - X_4 - T/C ↓ S/A, based upon sequence comparisons of genotypes 1 and 2 (ref. 27). Other genotypes including type 5 show a number of exceptions to this consensus at the P1′ and P6 positions (Fig 2A). For example, substitutions of asparagine (type 3b), aspartatic acid (types 10a and 5a) and glycine (type 11a) are found at the P1′ site at the NS4B/NS5A junction. The P6 site between NS5A and NS5B shows even greater variability, where many sequences do not contain an

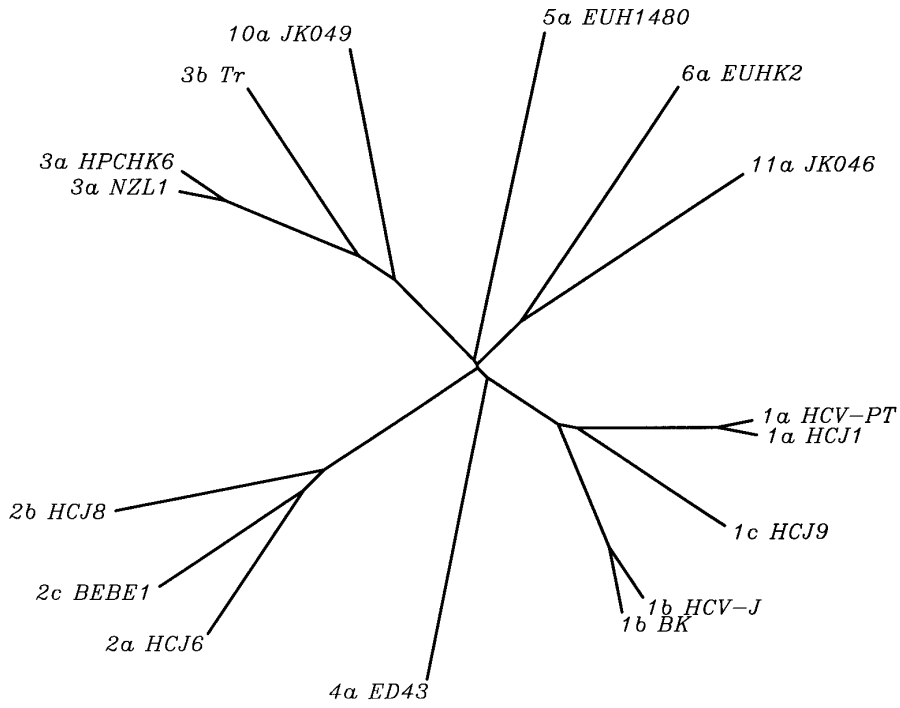

FIG. 3. Phylogenetic comparison of HCV genotype 5a sequence (5′NCR and complete coding region) with representatives of the other genotypes, shown as an unrooted tree using the program DNADIST (maximum likelihood distances) and NEIGHBOR in the PHYLIP package (25).

acidic residue, although in these cases (types 3 and 11a), one is found at P7. Mutagenesis studies have shown considerable tolerance for substitutions at the four sites required for proteolytic cleavage by NS3, particularly at the *cis* site between NS3 and NS4A (28, 29). None of the amino acid changes observed between genotypes would be expected to alter processing efficiency in such *in vitro* assays. Overall, therefore, the targets for both the NS2-3 and the NS3 protease appear to be subject to constraints on sequence variability which are not reproduced by *in vitro* assay.

Variability in the putative interferon sensitivity determining region in NS5A (Fig 2B), has been shown, at least for type 1b infections, to correlate with response to α or β interferon treatment (30). Variants of type 1b, with greater than four amino acid substitutions compared with HCV-J, show a substantially higher response rate to interferon treatment, although it remains unclear whether a similar differentiation can be made between variants of other genotypes. The two genotypes associated with a favourable outcome of treatment (types 2 and 3) show numerous differences in this region. However while type 4 has previously been shown to respond poorly to treatment (31) the sequence of ED43 (ref. 22) shows 18 differences from type 1b. There is little information on the response to treatment of type 5 and 6, although interestingly jk046 shows only three amino acid differences from HCV-J while the other type 6 variant, EUHK2, shows 15 differences.

The hypervariable region of E2 shows considerable variability both between and within genotypes (Fig 2C). Only two amino acid residues are completely conserved between sequences of all six genotypes (threonine at position 385 and glycine at position 406). Previous sequence comparisons of genotypes 1 and 2 also identified a conserved glycine residue at position 389, although this is not conserved in type 6. A third position (amino acid 403) shows either a phenylalanine or leucine residue conserved in all 6 genotypes. All other positions contain numerous alternative amino acids, with no obvious evidence for conservation of charge or size of residue.

The E1 and E2 proteins of type 5 showed a similar distribution of potential N-linked glycosylation sites as other genotypes. The E1 protein contained four sites (positions 196, 209, 234 and 305; numbered as in the HCV-PT polyprotein sequence, ref 32) which are conserved amongst all variants. Glycosylation in the E2 protein was somewhat more variable. The sequence of type 5 predicted nine sites conserved amongst all other genotypes (positions 417, 423, 430, 448, 532, 556, 576, 623 and 634), but lacked the site at position 476 present in all genotypes except types 1b, 2c and 3b. It also lacked the more variable sites at positions 524 (type 6a and jk046), 540 (types 1, 2 and 4) and 708 (type 6a only).

Phylogenetic analysis of complete genomic sequences was carried out using the sequence of EUH1480 and representative sequences of the other genotypes (Fig 3). The type 5a sequence branched from other clades at a point close to the common origin of clades 1 (and 4), 2, 3 and 6. A separate bootstrap analysis using the programs SEQBOOT and CONSENSE in the PHYLIP package (25) showed no consistent shared branching of type 5 with any other genotype. The closest association was with sequences in the type 3 clade, where 67% of trees contained a common ancestor for both, while 58% of trees contained a common ancestor for clades 3, 5 and 6. Neither of these values were sufficiently robust to indicate a significantly greater evolutionary relatedness of type 5 to these rather than to other genotypes. Robust grouping of type 5 with other genotypes was similarly not observed in a parallel analysis of inferred amino acid sequences of complete coding sequences (data not shown). The bootstrap re-sampling results are consistent with the approximately equal degree of nucleotide and amino acid sequence divergence observed between type 5 and each other genotype.

With the determination of the complete genomic sequence from isolate EUH1480, genotype 5 can be confirmed as a distinct genotype. Complete genomic sequences from each of the main genotypes clarify the classification system of HCV and will aid in future diagnosis and classification of particular isolates. The availability of cDNA clones generated in this study together with those from previous studies (16,22) will also enable comparisons between genotypes to be extended by examining protein expression, processing and function. Such studies will be initiated with the long term goals of developing inhibitors with broad spectrum activity against all HCV types.

ACKNOWLEDGMENTS

This work was supported by an MRC project grant to R. M. Elliott, E. A. McCruden and P. Simmonds.

REFERENCES

1. Simmonds, P., Alberti. A., Alter, H. J., Bonino, F., Bradley, D. W., Brechot. C., Brouwer. J. T., Chan. S-W., Chayama, K., Chen, D-S., Choo, Q-L., Colombo, M., Cuypers, H. T. M., Date, T., Dusheiko, G. M., Esteban, J. I., Fay, O., Hadziyannis, S. J., Han, J., Hatzakis, A., Holmes, E. C., Hotta, H., Houghton, M., Irvine, B., Kohara, M., Kolberg, J. A., Kuo, G., Lau, J. Y. N., Lelie, P. N., Maertens, G., McOrmish, F., Miyamura, T., Mizokami, M., Nomoto, A., Prince, A. M., Reesink, H. W., Rice, C., Roggendorf, M., Schalm, S. W., Shikata, T., Shimotohno, K., Stuyver, L., Trepo, C., Weiner, A., Yap, P. L., and Urdea, M. S. (1994) *Hepatology* **19,** 1321–1324.
2. Simmonds, P. (1995) *Hepatology* **21,** 570–583.
3. Bukh, J., Purcell, R. H., and Miller, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2276–2279.
4. Simmonds, P., McOrmish, F., Yap, P. L., Chan, S.-W., Lin, C. K., Dusheiko, G., Saeed, A. A., and Holmes, E. C. (1993) *J. Gen. Virol.* **74,** 661–668.
5. Bukh, J., Purcell, R. H., and Miller, R. H (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8239–8243.
6. Simmonds, P., Smith, D. B., McOrmish, F., Yap, P. L., Kolberg, J., Urdea, M. S., and Holmes, E. C. (1994) *J. Gen. Virol.* **75,** 1053–1061.
7. Bukh, J., Purcell, R. H., and Miller, R. H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8234–8238.
8. Stuyver, L., van Arnhem, W., Wyseur, A., Hernandez, F., Delaporte, E., and Maertens, G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10134–10138.
9. Simmonds, P., Holmes, E. C., Cha, T.-A., Chan, McOrmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J., and Urdea, M. S. (1993). *J. Gen. Virol.* **74,** 2391–2399.
10. Tokita, H., Okamoto, H., Tsuda, F., Song, P., Nakata, S., Chosa, T., Iizuka, H., Mishiro, S., Miyakawa, Y., and Mayumi, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11022–11026.
11. Tokita, H., Okamoto, H., Luengrojanakul, P., Vareesangthip, K., Chainuvati, T., Iizuka, H., Miyakawa, Y., and Mayumi, M. (1995) *J. Gen. Virol.* **76,** 2329–2335.
12. Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F., Lesmana, L. A., Song, P., Miyakawa, Y., and Mayumi, M. (1996) *J. Gen. Virol.* **77,** 293–301.
13. Mizokami, M., Gojobori, T., Ohba, K. I., Ikeo, K., Ge X. M., Ohno, T., Orito, E., and Lau, J. Y. N. (1996) *J. Hepatol.* **24,** 622–624.
14. Simmonds, P., Mellor, J., Holmes, E. C., Tampresert, S., Sakuldamrongpanich, T., Nuchaprayoon, C., and Smith, D. B. (1996) *J. Gen. Virol.* **77,** 3013–3024.
15. Delambellerie, X., Charrel, R. N., Attoui, H., and Demicco, P. (1997) *J. Gen. Virol.* **78,** 45–51.
16. Adams, N. J., Chamberlain, R. W., Taylor, L. A., Davidson, F., Lin, C. K., Elliott, R. M., and Simmonds, P. (1997) *Biochem. Biophys. Res. Comm.* (in press).
17. Davidson F., Simmonds, P., Ferguson, J. C., Jarvis, L. M., Dow, B. C., Follett, E. A. C., Seed, C. R. G., Krusius, T., Lin, C., Medgyesi., G, A., Kiyokawa, H., Olim, G., Duraisamy, G., Cuypers, T., Saeed, A. A., Teo, D., Conradie, J., Kew, M. C., Lin, M., Nuchaprayoon, C., Ndimbie, O. K., and Yap, P. L. (1995) *J. Gen. Virol.* **76,** 1197–1204.
18. Cha T. A., Beall, E., Irvine, B., Kolberg, J., Chien, D., Kuo, G., and Urdea, M. S. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7144–7148.
19. McOrmish, F., Yap, P. L., Dow, B. C., Follett, E. A. C., Seed, C. R. G., Keller, A. J., Cobain, T. J., Krusius, T., Kolho, E., Naukkarinen, R., Lin, C., Lai, C., Leong, S., Medgyesi., G. A., Hejjas, M., Kiyokawa, H., Fukada, K., Cuypers, T., Saeed, A. A., Alrasheed, A. M., Lin, M., and Simmonds, P. (1994) *J. Clin. Micro.* **32,** 882–892.
20. Prescott, L. E., Simmonds, P., Lai, C. L., Chan, N. K., Pike, I., Yap, P. L., and Lin, C. K. (1996) *J. Med. Virol.* **50,** 168–175.
21. Mellor J., Walsh, E. A., Prescott, L. E., Jarvis, L. M., Davidson, F., Yap, P. L., Simmonds, P., Nowicki, M. J., Mosley, J. W., Lin, C. K., Lai, C. L., deOlim, G., Martins, I. A., Ong, Y. W., Teo, D., Lin, M., Nuchprayoon, C., and Tansprasert, S. (1996) *J. Clin. Micro.* **34,** 417–423.
22. Chamberlain, R. W., Adams, N., Saeed, A. A., Simmonds, P., and Elliott, R. M. (1997) *J. Gen. Virol.* (in press).
23. Tellier, R., Bukh, J., Emerson, S. U., Miller, R. H., and Purcell, R. H. (1996) *J. Clin. Microbiol.* **34,** 3085–3091.
24. Staden, R. (1987) Nucleic acid and protein analysis: A practical approach. Edited by M. J. Bishop and C. J. Rawlings: IRL Press Oxford.

25. Felsenstein, J. (1993) PHYLIP Inference Package version 3.5, Seattle: Department of Genetics, University of Washington.
26. Reed, K. E., Graouki, A., and Rice, C. M. (1995) *J. Virol.* **69,** 4127–4136.
27. Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M., and Rice, C. M. (1993) *J. Virol.* **67,** 2832–2843.
28. Bartenschlager, R., Ahlbornelaake, L., Yasargel, K., Mous, J., and Jacobsen, H. (1995) *J. Virol.* **69,** 198–205.
29. Kolykhalov, A. A., Agapov, E. V., and Rice, C. M. (1994) *J. Virol.* **68,** 7525–7533.
30. Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C., Ogura, Y., Izumi, N., Marumo, F., and Sato, C. (1996) *N. Eng. J. Med.* **334,** 77–81.
31. El Zayadi, A., Simmonds, P., Dabbous, H., Prescott, L., Selim, O., and Ahdy, A. (1996) *J. Viral. Hep.* **3,** 261–264.
32. Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina Selby, R., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G., and Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2451–2455.