Evolutionary analysis of the 5′-terminal region of hepatitis G virus isolated from different regions in China

Ping An,[1,3] Li Wei,[3] Xueyun Wu,[2] Naoya Yuhki,[1] Stephen J. O'Brien[1] and Cheryl Winkler[2]

[1,2] Laboratory of Genomic Diversity, National Cancer Institute[1] and Intramural Research Support Program[2], SAIC Frederick, National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, MD 21702-1201, USA

[3] The First Teaching Hospital, Lanzhou Medical College, Lanzhou 730000, People's Republic of China

We have determined the nucleotide sequence of the 5′-terminal region of the hepatitis G virus (HGV) genome in 11 hepatitis patients from three cities in China. Phylogenetic analyses revealed that the Chinese isolates were genetically distinct from previously described West African isolates (type 1) and American, European and East African isolates (type 2), with a mean sequence divergence of approximately 10%. The mean divergence between isolates from Lanzhou, in the northwest of China, and those from Shanghai and Nanjing, on the east coast of China, was 5% (range 3–7%). The isolates from Shanghai and Nanjing were closely related to a common strain in Japan, while some of those from Lanzhou were closely related to a southeast Asian type 3 isolate. Thus, the Chinese isolates belong to the type 3 variant of HGV.

Recently, two new viruses of humans were independently discovered and named GB virus C (GBV-C) and hepatitis G virus (HGV), respectively. A comparison of the full-length sequences revealed that GBV-C and HGV represent different strains of the same species of virus. HGV/GBV-C, referred to here as HGV, have 64% amino acid similarity with GB virus A (GBV-A), 29% with hepatitis C virus (HCV) and 28% with GB virus B (GBV-B). These viruses represent discrete virus lineages within the family *Flaviviridae*, based upon genomic organization and phylogenetic analysis (Simons *et al.*, 1995; Linnen *et al.*, 1996; Leary *et al.*, 1996).

HGV infections have been detected in 1–2% of blood donors in many countries and in 14–52% of patients with various types of viral hepatitis (Simons *et al.*, 1995; Linnen *et al.*, 1996; Miyakawa *et al.*, 1997; M. J. Alter *et al.*, 1997). The role of HGV infection in liver diseases is not well understood. It appears that persistent infections with HGV are common, but they do not seem to be associated with significant hepatic injuries (Masuko *et al.*, 1996; M. J. Alter *et al.*, 1997; H. J. Alter *et al.*, 1997). HGV has also been found in 9–50% of patients with non-A–E acute or fulminant hepatitis (FH), but its causal role has not been established (Yoshiba *et al.*, 1995; Heringlake *et al.*, 1996; M. J. Alter *et al.*, 1997).

The genome of HGV consists of a positive-stranded RNA of approximately 9·5 kb with a single long open reading frame plus extended untranslated regions at the 5′ and 3′ ends (Leary *et al.*, 1996). Sequence analysis of currently available isolates showed that the 5′-terminal region of HGV is less highly conserved than that of HCV, although it contains several blocks of highly conserved sequences. Subsequently, phylogenetic analysis of 42 unique sequences of the 5′-terminal region of HGV isolates worldwide classified HGV into three types, based on an evolutionary distance of approximately 0·1 and five subtypes (1a, 1b, 2a, 2b and 3) separated by a distance of 0·034–0·056. Type 1 isolates were found exclusively in West Africa and type 2 in the USA and Europe, while type 3 (two isolates) was unique to southeast Asia (Muerhoff *et al.*, 1996*a*). Okamoto *et al.* (1997) analysed the entire nucleotide sequences of the five HGV isolates and tentatively classified them into three genotypes with an evolutionary distance $> 0{\cdot}13$.

In this study, we have determined nucleotide sequences at the 5′-terminal region of HGV from 11 patients in three cities in China. These sequences were then compared by phylogenetic analysis to previously reported representative sequences. Sera from 11 patients infected with HGV were used as the source of viral RNA. Six serum samples were collected from the city of Lanzhou (LZ), in the northwest of China, three were from Shanghai (SH) and two were from Nanjing (NJ), on the east coast of China. Subjects LZ1 and LZ5 were coinfected with HCV and HBV and diagnosed with chronic persistent hepatitis (CPH) and chronic active hepatitis (CAH), respectively; LZ3 and LZ4 were diagnosed with CPH with HCV infection; LZ2, LZ6, NJ1, NJ2, SH1, SH2 and SH3 had non-

Author for correspondence: Cheryl Winkler. Postal address: Bldg 560, Room 11-85, NCI–FCRDC, Frederick, MD 21702, USA. Fax +1 301 846 1909. e-mail winkler@ncifcrf.gov

The nucleotide sequences reported in this paper have been submitted to GenBank and assigned the accession numbers U89148–U89158.

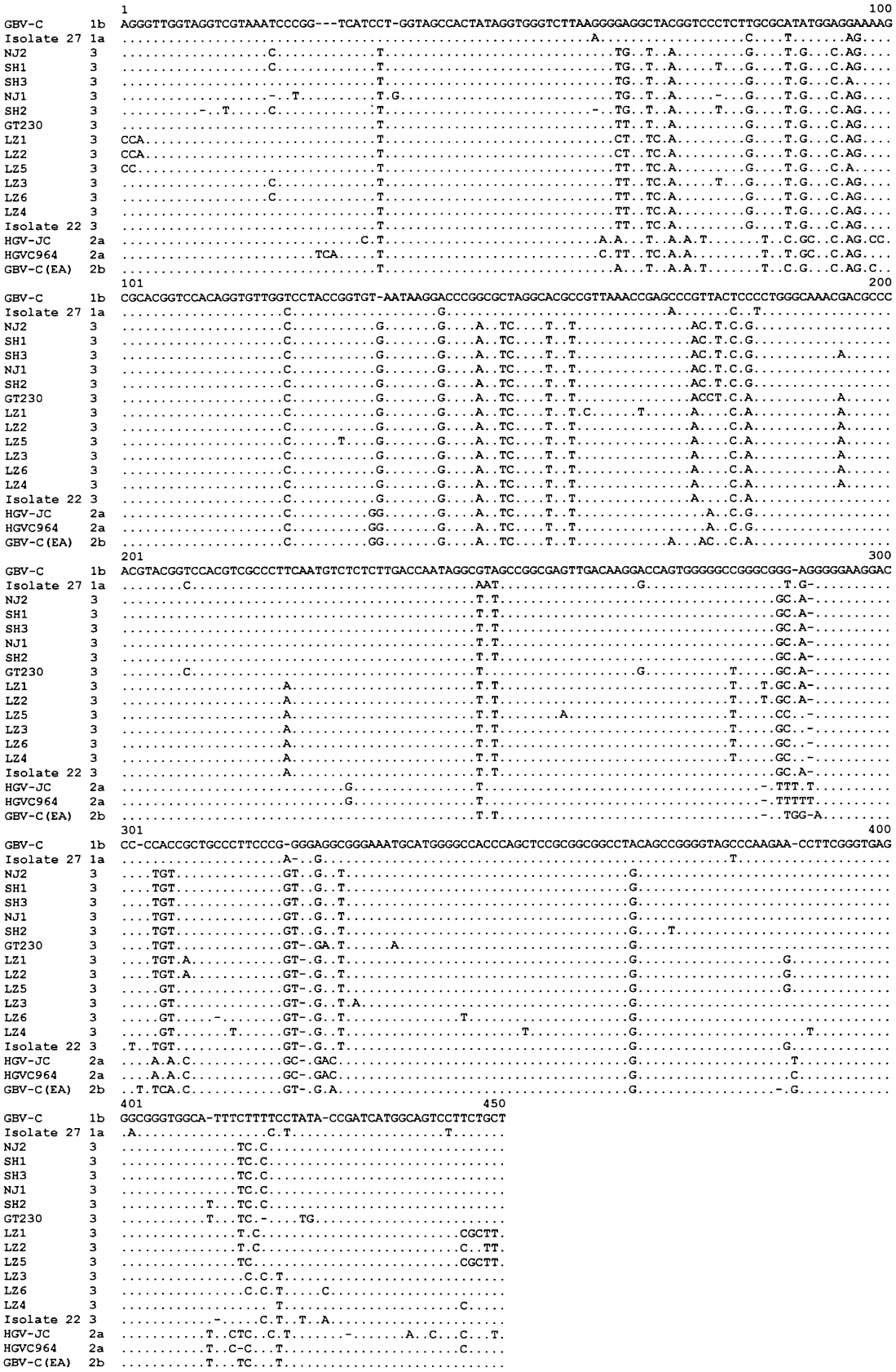

Fig. 1. For legend see facing page.

A–E hepatitis and were diagnosed with CPH, liver cirrhosis, acute hepatitis (AH), AH, FH, CPH and CAH, respectively.

Viral RNA was extracted from 140 µl serum using a QIAamp HCV kit (Qiagen). The extracted RNA was reverse-transcribed to cDNA with SuperScript II RNase H reverse transcriptase and random hexamers (Gibco BRL). HGV sequences from the 5′-terminal region (positions 112–548, numbering according to Leary *et al*., 1996) were amplified by RT–PCR, using the published S2/5R primer pair (Muerhoff *et al*., 1996*b*). Amplified products (438–440 bp) were cloned into the PCRII vector using the TA cloning system (Invitrogen). Subcloned DNA was purified using Qiagen columns. Two or three clones from each individual were sequenced in both directions on an ABI model 373 DNA sequencer using the dye terminator cycle sequencing kit (Perkin-Elmer). A consensus sequence was then generated for each individual. In all of the isolates from Lanzhou, only minor base substitutions were observed, with 99·7% nucleotide sequence identities among three clones from the same patient, while all isolates from Shanghai and Nanjing were identical between two clones from the same patient. For isolates LZ1, LZ2, LZ4, LZ5, SH1 and SH3, a longer sequence fragment (535 or 586 bp) which flanked the above region was amplified with other primer pairs (Muerhoff *et al*., 1996*b*) (sequences not shown) to verify the 5′ and 3′ end sequences presented here.

Nucleotide sequences were aligned using PILEUP from the GCG package. Evolutionary distances between sequences were estimated using the DNAML model of the DNADIST program in order to correct for different rates of transitions and transversions and different frequencies of nucleotides (Felsenstein, 1993). Phylogenetic trees were constructed by the maximum likelihood and neighbour-joining (N-J) methods of the PHYLIP package and by the maximum parsimony method of PAUP (phylogenetic analysis using parsimony) (Swofford, 1993). Both N-J and PAUP trees were evaluated using 100 bootstrap iterations to provide an index of confidence for each topological node. Bootstrap values $> 70\%$ were considered strong support for the adjacent node (Hillis & Bull, 1993).

Sixteen previously published sequences of HGV isolates were used in the phylogenetic analysis: GBV-C (GenBank accession no. U36380) (Leary *et al*., 1996); GBV-C(EA) (U63715) (Erker *et al*., 1996); HGV PNF2161 (U44402), HGV-JC (U45966) (Linnen *et al*., 1996); HGV-Iw (U87255) (Shao *et al*., 1996); GT110 (D90600) and GT230 (D90601) (Okamoto *et al*., 1997); HGVC964 (U75356) (Zhou *et al*., 1996); isolate 12 (U59529), isolate 22 (U59539), isolate 24 (U59541), isolate 26 (U59543), isolate 27 (U59544), isolate 28 (U59545), isolate 34 (U59550) and isolate 39 (U59555) (Muerhoff *et al*., 1996*a*).

The 5′-terminal sequences from 11 Chinese isolates and 16 previously reported isolates were aligned for genetic analyses and representative sequences are presented in Fig. 1. The nucleotide sequence identities of the 5′-terminal region of the Chinese isolates reported here with eight entire viral genomes available to date (HGV PNF2161, HGV-JC, HGV-Iw, GBV-C and GBV-C(EA), HGVC964, GT110 and GT230) were 88·0–90·5%, 87·8–89·8%, 88·7–91·2%, 88·0–91·4%, 90·7–92·8%, 89·6–91·9%, 87·8–90·5% and 92·5–96·6%, respectively. The nucleotide sequence identities with the southeast Asian isolate 22 were 94·8–97·0%. These data indicated that the Chinese isolates were most closely related to the southeast Asian isolate and Japanese isolate GT230, but distantly related to the USA, European, and west and east African isolates. Among Chinese isolates, there were sequence identities of 93·0–99·8%, indicating a moderate degree of genetic heterogeneity.

To determine the phylogenetic relationship between the HGV isolates from China and elsewhere, we used three different phylogenetic methods. This increased the reliability of the derived topologies since the tree-building algorithms rely on different assumptions (Felsenstein, 1988; Swofford *et al*., 1996). The topology of the trees was consistent for all methods applied, thus only the N-J tree is presented (Fig. 2). These analyses revealed that the variability of HGV was structured into three divergent main groups of sequences with very strong bootstrap value support by N-J (93–100%) and PAUP (84–100%). These three major groups corresponded to previously defined HGV types 1, 2 and 3 (Muerhoff *et al*., 1996*a*) or genotypes G1, G2 and G3 (Okamoto *et al*., 1997). Each of the major groups contained two subgroups. Four of these subgroups corresponded to the previously defined subtypes 1a, 1b, 2a and 2b, respectively (Muerhoff *et al*., 1996*a*). The isolates from Lanzhou fell into the same subgroup as the southeast Asian isolate 22, a representative of type 3 (Muerhoff *et al*., 1996*a*). The isolates from Shanghai and Nanjing clustered with the Japanese isolate GT230, which represents prevalent strains in Japan and genotype G3 (Okamoto *et al*., 1997). The Chinese isolates reported here thus belong to HGV type 3. The mean distance of type 3 to representative examples of type 1 was 0·106 (range 0·078–0·140) and to those of type 2 was 0·090 (range 0·063–0·124). These distance values were comparable to the previously reported mean distance value of 0·114 (range 0·051–0·159) between major types (Muerhoff *et al*., 1996*a*). Judging from the distance values and the branch lengths of the tree, type 3 isolates were genetically more closely related to type 2 isolates than to type 1 isolates.

Fig. 1. The alignment of nucleotide sequences of the 5′-terminal region from 11 new Chinese HGV isolates and seven previously reported isolates. The type/subtype is shown to the right of each isolate. The sequences are compared to the top GBV-C sequence (Leary *et al*., 1996). Dashes represent gaps introduced to optimize the alignment and thus 450 nucleotide sites were obtained. Identical nucleotides are shown as solid dots and nucleotide substitutions are indicated.

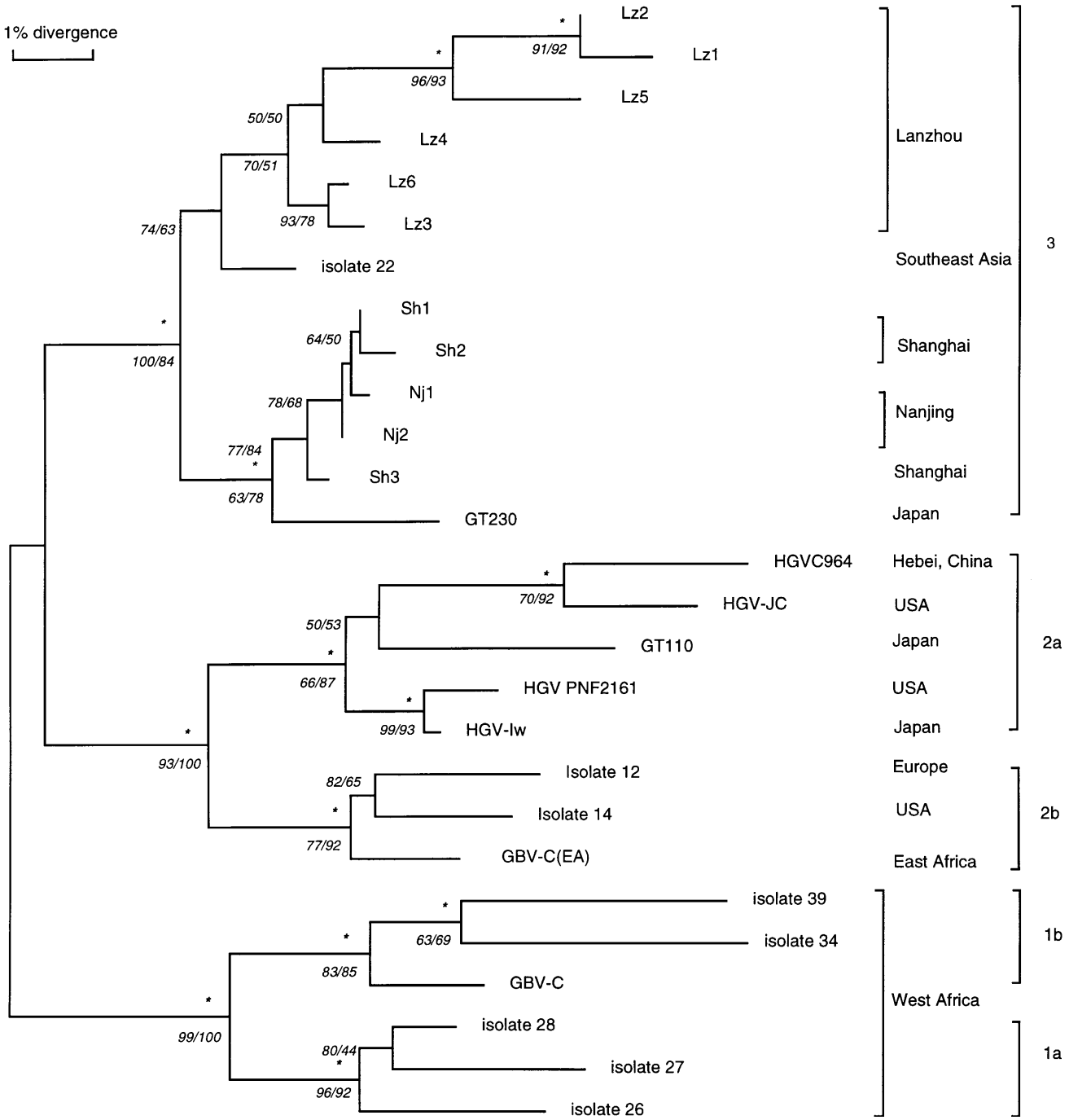

Fig. 2. Neighbour-joining tree derived from the 5′-terminal region of HGV from 11 Chinese isolates and 16 previously reported isolates (see text). The tree is drawn at midpoint. Horizontal lengths are drawn to scale; vertical separations are for clarity only. Asterisks indicate significant support ($P < 0{\cdot}01$) for major adjacent node by the maximum likelihood method. Numbers in italics (e.g. *91/92*) are bootstrap values of the form neighbour-joining/PAUP in support of the adjacent node; only one of those values, greater than 50%, is shown. The HGV types/subtypes and geographical origins of isolates are indicated by brackets on the right.

The genetic distance within and between HGV types/subtypes is listed in Table 1. Although the mean distances between subtypes (1a and 1b, 2a and 2b) were only a little greater than the mean distances within subtypes, and ranges of distances between and within subtypes also overlapped to some extent, the division of types (type 1 and type 2) into subtypes was supported by moderate to high bootstrap values (66–96%) and the maximum likelihood analysis (Fig. 2). Muerhoff *et al.* (1996 *a*) also found that the distances of HGV isolates within subtypes (1a, 1b, 2a and 2b) ranged from 0·00

Table 1. Means and ranges of nucleotide distance within and between phylogenetic groups derived from the 5′-terminal region of HGV

Ranges of nucleotide distances are given in parentheses.

HGV subtypes	No.*	1a	1b	2a	2b	3†
1a	3	0·0392 (0·0328–0·0494)				
1b	3	0·0797 (0·0571–0·1060)	0·0631 (0·0598–0·0697)			
2a	4	0·1201 (0·1077–0·1385)	0·1181 (0·1017–0·1410)	0·0520 (0·0120–0·0725)		
2b	3	0·1098 (0·1018–0·1264)	0·1162 (0·1022–0·1299)	0·0681 (0·0403–0·0867)	0·0358 (0·0349–0·0374)	
3	13	0·1008 (0·0781–0·1366)	0·1109 (0·0854–0·1395)	0·0940 (0·0729–0·1239)	0·0829 (0·0627–0·1101)	0·0378 (0·0023–0·0698)

* Number of sequences compared.
† The low distance value is affected by the low degree of diversity of isolates from Shanghai and Nanjing.

to 0·083 (mean 0·046), while the distance between subtypes ranged from 0·065 to 0·092. Therefore, it is currently impossible to specify a cut-off value for defining HGV types and subtypes. It is to be hoped that analysis based on a large number of entire HGV genomes would elucidate this issue.

Type 3 isolates were further segregated into two clusters with moderate bootstrap value support by N-J (63–74%) and PAUP (63–78%). The mean distance value between these two subgroups was 0·047. The mean distance value between isolates from Lanzhou and those from Shanghai and Nanjing was 0·048, ranging from 0·030 (LZ3 vs SH1) to 0·070 (LZ1 vs SH2). The isolates from Shanghai and those from Nanjing formed a tight cluster. These data correlated well with their geographical origin, since Lanzhou is approximately 1500 km away from Shanghai while the latter is only 250 km away from Nanjing. The mean distance of Japanese isolate GT230 to the isolates from Shanghai and Nanjing and to those from Lanzhou was 0·032 (range 0·028–0·038) and 0·054 (range 0·043–0·068), respectively. The southeast Asian isolate 22 was most closely related to LZ3, LZ6 and LZ4, with distance values of 0·027–0·028.

Much lower genetic diversity of 0·015 (range 0·002–0·038) was observed among the isolates from Shanghai and Nanjing than among those from Lanzhou (mean 0·033, range 0·009–0·057). The low degree of diversity among isolates from Shanghai and Nanjing possibly reflected a more recent spread of HGV variants in this area. Careful scrutinizing of the histories of these patients did not provide evidence for epidemiological linkage (i.e. no transfusion histories or physical contacts) among these patients. On the other hand, the relatively high degree of diversity among isolates from Lanzhou might suggest the long-term existence of the HGV variants in this population.

To our surprise, HGVC964, from a blood donor in north China (Zhou *et al*., 1996), appeared to be a recombinant between type 3 and subtype 2a strains. Phylogenetic analyses using only the 5′-terminal region placed this isolate with HGV-JC in the subtype 2a clade (Fig. 2). Indeed, HGVC964 shared more common nucleotide characters with HGV-JC than with other isolates (Fig. 1). The genetic distance of HGVC964 to HGV-JC was only 0·040, while to all type 3 isolates it was 0·073–0·099. We therefore constructed trees (not shown) based on three different regions of the HGV genome using eight full-length isolates. We found that the 5′-terminal region of HGVC964 most closely resembled HGV-JC (bootstrap value 62–74%), with distances of 0·048 to HGV-JC and 0·073 to GT230. In contrast, the E1–E2 and E1–NS5b regions were most closely related to the type 3 isolate GT230 (bootstrap values 92–100%), with distances of 0·12 (for both regions) to GT230 and 0·14 and 0·16 to HGV-JC, respectively.

We are very grateful to Drs Claiborne Stephens and Jill Slattery for expert assistance in phylogenetic analyses and to Dr George Nelson for helpful discussion of the manuscript. Dr Ping An was supported in part by a WHO fellowship. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

Alter, H. J., Nakatsuji, Y., Melpolder, J., Wages, J., Wesley, R., Shih, J. W. & Kim, J. P. (1997). The incidence of transfusion-associated hepatitis G virus infection and its relation to liver disease. *New England Journal of Medicine* **336**, 747–754.

Alter, M. J., Gallagher, M., Morris, T. T., Moyer, L. A., Meeks, E. L., Krawczynski, K., Kim, J. P. & Margolis, H. S. (1997). Acute non-A–E hepatitis in the United States and the role of hepatitis G virus infection. *New England Journal of Medicine* **336**, 741–746.

Erker, J. C., Simons, J. N., Muerhoff, A. S., Leary, T. P., Chalmers, M. L., Desai, S. M. & Mushahwar, I. K. (1996). Molecular cloning and characterization of a GB virus C isolate from a patient with non-A–E hepatitis. *Journal of General Virology* **77**, 2713–2720.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* **22**, 521–565.

Felsenstein, J. (1993). PHYLIP: phylogenetic inference package, version 3.5c. University of Washington, Seattle, Wash., USA.

Heringlake, S., Osterkamp, S., Trautwein, C., Tillmann, H., Boker, K., Muerhoff, S., Mushahwar, I. K., Hunsmann, G. & Manns, M. (1996). Association between fulminant hepatic failure and a strain of GBV virus C. *Lancet* **348**, 1626–1629.

Hillis, D. M. & Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**, 182–192.

Leary, T. P., Muerhoff, A. S., Simons, J. N., Pilot-Matias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996). Sequence and genomic organization of GBV-C: a novel member of the *Flaviviridae* associated with human non-A–E hepatitis. *Journal of Medical Virology* **48**, 60–67.

Linnen, J., Wages, J., Jr., Zhang-Keck, Z. Y., Fry, K. E., Kawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, P., Fung, K., Nakatsuji, Y., Shih, J. W., Young, L. M. P., Jr, Hoover, C., Fernandez, J., Chen, S., Zou, J.-C., Morris, T., Hyams, K. C., Ismay, S., Lifson, J. D., Hess, G., Foung, S. K. H., Thomas, H., Bradley, D., Margolis, H. & Kim, J. P. (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**, 505–508.

Masuko, K., Mitsui, T., Iwano, K., Yamazaki, C., Okuda, K., Meguro, T., Murayama, N., Inoue, T., Tsuda, F., Okamoto, H., Miyakawa, Y. & Mayumi, M. (1996). Infection with hepatitis GB virus C in patients on maintenance hemodialysis. *New England Journal of Medicine* **334**, 1485–1490.

Miyakawa, Y. & Mayumi, M. (1997). Hepatitis G virus –a true hepatitis virus or an accidental tourist? [editorial]. *New England Journal of Medicine* **336**, 795–796.

Muerhoff, A. S., Simons, J. N., Leary, T. P., Erker, J. C., Chalmers, M. L., Pilot-Matias, T. J., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996*a*). Sequence heterogeneity within the 5′-terminal region of the hepatitis GB virus C genome and evidence for genotypes. *Journal of Hepatology* **25**, 379–384.

Muerhoff, A. S., Simmons, J. N., Erker, J. C., Desai, S. M. & Mushahwar, I. K. (1996*b*). Identification of conserved nucleotide sequences within the GB virus C 5′-untranslated region: design of PCR primers for detection of viral RNA. *Journal of Virological Methods* **62**, 55–62.

Okamoto, H., Nakao, H., Inoue, T., Fukuda, M., Kishimoto, J., Lizuka, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1997). The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from Japan. *Journal of General Virology* **78**, 737–745.

Shao, L., Shinzawa, H., Shikawa, K., Zhang, X., Ishbashi, M., Misawa, H., Yamada, N., Togashi, H. & Takahashi, T. (1996). Sequence of hepatitis G virus genome isolated from a Japanese patient with non-A–E-hepatitis: amplification and cloning by long reverse transcription-PCR. *Biochemical and Biophysical Research Communications* **228**, 785–791.

Simons, J. N., Leary, T. P., Dawson, G. J., Pilot-Matias, T. J., Muerhoff, A. S., Schlauder, G., Desai, S. M. & Mushahwar, I. K. (1995). Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine* **1**, 564–569.

Swofford, D. L. (1993). PAUP: phylogenetic analysis using parsimony, version 3.1. Champaign: Illinois Natural History Survey.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. (1996). Phylogenetic inference. In *Molecular Systematics*, 2nd edn, pp. 407–514. Edited by D. M. Hillis, C. Moritz & B. K. Mable. Sunderland, Mass.: Sinauer Associates.

Yoshiba, M., Okamoto, H. & Mishiro, S. (1995). Detection of the GBV-C hepatitis virus genome in serum from patients with fulminant hepatitis of unknown aetiology. *Lancet* **346**, 1131–1132.

Zhou, Y. S., Chen, W., Zhao, Q. M., Zhao, H. L., Zhang, J. S., Xu, J. S. & Wang, H. T. (1996). cDNA cloning and sequencing of HGV genome from Chinese. *Bulletin of Academy of Military Medical Sciences* **20**, 249–253.

Received 10 March 1997; Accepted 29 May 1997