African origin of GB virus C/hepatitis G virus

Yasuhito Tanaka[a], Masashi Mizokami[a,*], Etsuro Orito[a], Ken-ichi Ohba[a], Takanobu Kato[a], Yutaka Kondo[a], Innocent Mboudjeka[b], Leopold Zekeng[b], Lazare Kaptue[b], Blaise Bikandou[c], Pierre M'Pele[c], Jun Takehisa[d], Masanori Hayami[d], Yoshiyuki Suzuki[e], Takashi Gojobori[e]

[a]*Second Department of Medicine, Nagoya City University Medical School, Kawasumi, Mizuho, Nagoya 467, Japan*
[b]*Laboratoire d'Hématologie et d'Immunologie, Centre Hospitalier Universitaire, Yaounde, Cameroon*
[c]*Cité Louis Pasteur-Laboratoire National de Santé Publique, Brazzaville, Congo*
[d]*Institute for Virus Research, Kyoto University, Kyoto, Japan*
[e]*Center for Information Biology, National Institute of Genetics, Mishima, Japan*

Received 8 December 1997; revised version received 19 January 1998

Abstract Ninety-four GB virus C/hepatitis G virus (GBV-C/HGV) RNA-positive serum samples were obtained from all over the world. We found that all 15 GBV-C/HGV isolates from the Pygmies and the Bantu in the Central African region had a 12-amino acid indel (i.e. insertion or deletion) in the non-structural protein (NS) 5A region. Phylogenetic analyses of the NS5A region, using GBV-A as an outgroup, showed that these 15 isolates had diverged from the common ancestor much earlier than the remaining isolates, indicating an African origin of GBV-C/HGV.

Key words: GB virus C; Hepatitis G virus; Africa; Pygmy; Indel; NS5A; Phylogenetic analysis

1. Introduction

GB virus C (GBV-C) and hepatitis G virus (HGV) were identified in human sera from members of the population of West Africa [1] and the USA [2], respectively. These two viruses show 100% amino acid identity in the non-structural protein (NS) 3 region [2] and are similar to GBV-A, on the basis of genomic organization and molecular evolutionary analysis [1]. Recently, a phylogenetic analysis of the GBV-C/HGV 5′-untranslated region (5′-UTR) sequences demonstrated the presence of three major genotypes, GB type (type 1), HG type (2), and Asian type (3) [3,4]. Although these three genotypes strongly correlate with the geographic distribution [4,5], the evolutionary origin of GBV-C/HGV remains unclear because the previous studies have never identified a root of the phylogenetic tree.

To elucidate the origin of GBV-C/HGV, the nucleotide sequences of GBV-C/ HGV were determined in a large number of serum samples from Cameroon and Congo in the Central African region, including the Pygmies and the Bantu, and other areas worldwide. Using these nucleotide sequence data, we constructed phylogenetic trees for the NS3 and NS5A regions of GBV-C/HGV. In this tree, we located the root by using GBV-A as an outgroup. Our results show that GBV-C/HGV might be of African origin.

*Corresponding author. Fax: (81) (52) 852-0849.
E-mail: mizokami@med.nagoya-cu.ac.jp

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases with the following accession numbers: AB010155–AB010202.

2. Materials and methods

2.1. Collection of serum samples and detection of GBV-C/HGV RNA

Serum samples were collected, after obtaining informed consent, from 390 subjects: 174 from the Baka Pygmy and 186 from the Bantu in Cameroon, and 30 subjects from the Aka Pygmy of Congo in the Central African region. Serum samples from all subjects were stored at −80°C until assayed. Serum RNA was extracted from 100 μl of serum using the Sepa Gene-RVR kit (Sanko, Tokyo), precipitated with isopropanol, and washed with ethanol. GBV-C/HGV RNA was detected by nested reverse transcription polymerase chain reaction (RT-PCR) with primers derived from the 5′-UTR as previously described [3].

Forty-six of 390 serum samples from the African populations were found to be positive for GBV-C/HGV RNA. Other serum samples were collected from several subjects from Australia [6], Japan, and the Jewish population in Uzbekistan [7] as the source of viral RNA in this study (Table 1). In total, 94 GBV-C/HGV RNA-positive samples were used for the molecular evolutionary analysis. They included samples from patients with liver diseases and apparently healthy subjects.

2.2. Detection of the GBV-C/HGV genotype

As previously reported [3], in all samples positive for GBV-C/HGV RNA by RT-PCR, the GBV-C/HGV genotype was detected by restriction fragment length polymorphism (RFLP) analysis in the 5′-UTR. The GBV-C/HGV isolates were clearly separated into three distinct phylogenetic clusters (or genotypes), named the GB type (type 1), the HG type (2), and the Asian type (3), by comparison of nucleotide sequences of the 5′-UTR [3,4,8].

2.3. PCR amplification of the NS5A sequences

PCR amplification of the NS5A region sequences (462 bp) was performed with 1 μl of cDNA in a volume of 50 μl. The RT-PCR was performed with sense primer G5Af2 (5′-CAAAGTGTTACCATTGAYGGGGA-3′) and antisense primer G5Ar3 (5′-TAGTACGGAAAGAGCCACGTTGAAGAC-3′) for 9 min at 96°C, followed by 35 cycles, each cycle consisting of denaturation for 1 min at 96°C, annealing for 45 s at 60°C, and extension for 1 min at 72°C.

2.4. PCR amplification of the NS3 sequences

PCR amplification of the NS3 region sequences (489 bp) [9] was performed with 1 μl of cDNA in a volume of 50 μl. The RT-PCR was performed with sense primer 3gf1 (5′-GGCCGTGGACCCAAGTCCCAACAGAYGC-3′) and antisense primer 3gr3 (5′-ATCTCACCAACGTCCAGCTTTGTCTCAAT-3′) for 9 min at 96°C, followed by 35 cycles, each cycle consisting of denaturation for 1 min at 96°C, annealing for 45 s at 55°C, and extension for 1 min at 72°C.

2.5. Determination of the nucleotide sequences

The amplicons were analyzed by electrophoresis on 3% agarose gels, stained with ethidium bromide, and observed under ultraviolet light. The specificity of amplification was confirmed by direct sequencing of the amplified products with a 373A DNA Sequencer (Applied Biosystems, Foster City, CA).

2.6. Molecular evolutionary analysis

Molecular evolutionary analyses including GBV-A [10] were per-

PII S0014-5793(98)00083-0

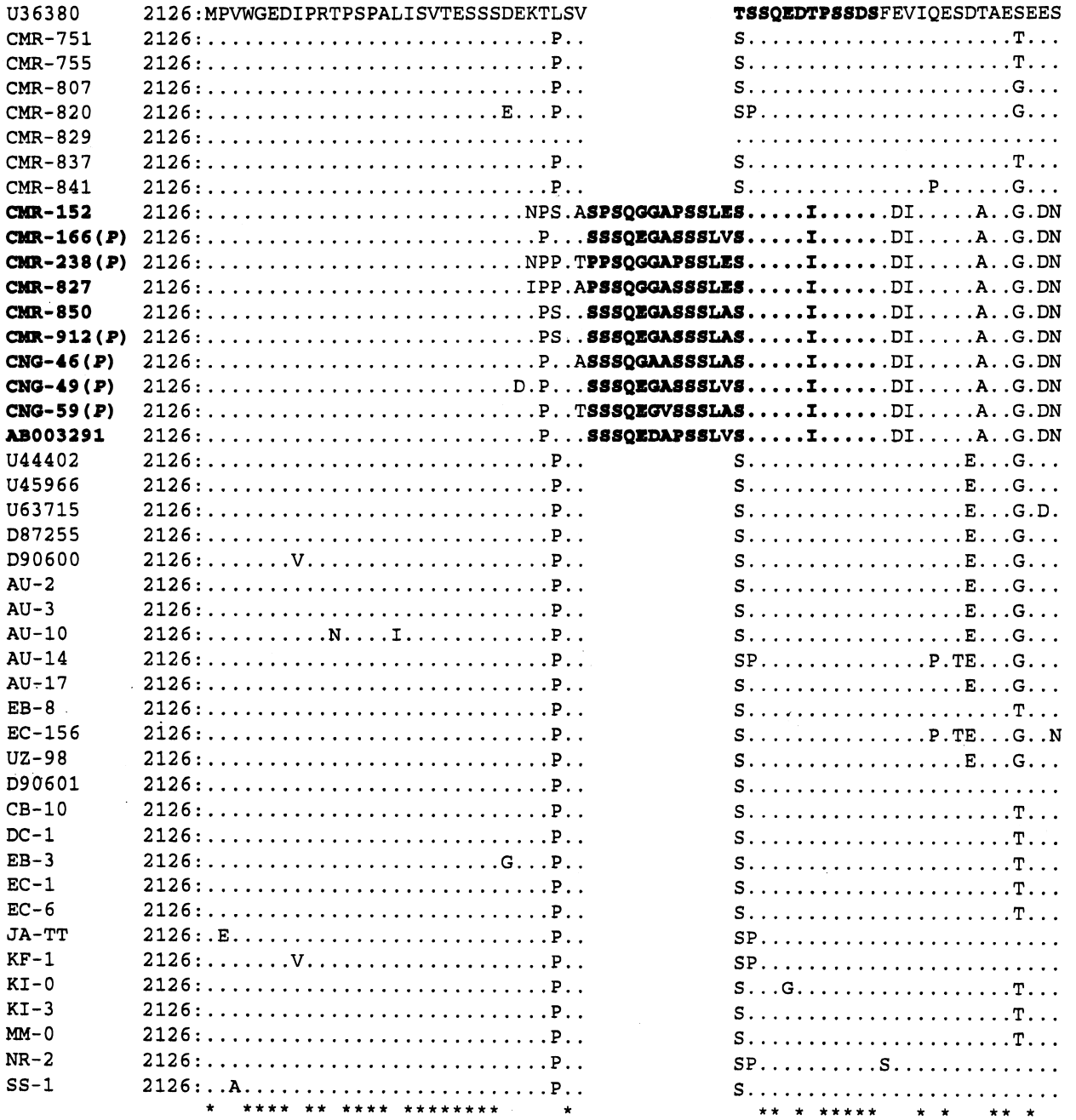

Table 1
Geographic distributions and 5′-UTR genotypes of GBV-C/HGV isolates from 94 subjects

Ethnic group		Number of isolates		
		GB type (type 1)	HG type (type 2)	Asian type (type 3)
Pygmies in Cameroon and the Congo	($n=20$)	18	2	0
Bantu in Cameroon	($n=26$)	26	0	0
Jewish in Uzbekistan	($n=14$)	2	12	0
Australian	($n=5$)	0	5	0
Japanese	($n=29$)	0	2	27
Total	($n=94$)	46	21	27

The three major GBV-C/HGV genotypes were designated the GB type, the HG type and the Asian type in our previous study [3], corresponding to genotypes 1, 2 and 3, respectively, as designated by Muerhoff et al. [4].

Fig. 1. Amino acid (aa) alignment of the GBV-C/HGV NS5A region. The isolates obtained in this study are represented as CMR, Cameroon; CNG, Congo; AU, Australia; UZ, Uzbekistan, followed by their identification numbers. EB-8, EC-156, CB-10, DC-1, EB-3, EC-1, JA-TT, KF-1, KI-0, KI-3, MM-0, NR-2 and SS-1 were isolated from Japan. U36380, U44402, U45966, U63715, D87255, D90600, D90601 and AB003291 were obtained from DDBJ/EMBL/GenBank DNA databases. Isolates obtained from the Pygmies are indicated as (P). Amino acids identical to those of GBV-C (U36380) are shown as a blot. Ten isolates, CMR-152, CMR-166, CMR-238, CMR-827, CMR-850, CMR-912, CNG-46, CNG-49, CNG-59 and AB003291, have a 12-aa indel and just downstream of this indel there are similar sequences shown in bold letters.
←

formed to elucidate the relationship between the isolates in this study and previously reported GBV-C/HGV strains [2,11–15]. Using the computer program ODEN version 1.1.1 [16], the number of nucleotide substitutions per site between all these isolates was estimated by the 6-parameter method [17]. Based on these values, phylogenetic trees were constructed by the neighbor-joining method [18]. To confirm the reliability of the phylogenetic tree, bootstrap resampling tests were performed 1000 times [19].

3. Results

3.1. High prevalence of GBV-C/HGV RNA in the Central African region

We found that the prevalence of GBV-C/HGV was quite high in the Central African region. In fact, of the 186 serum samples from the Bantu in Cameroon, 26 (14.0%) were positive for GBV-C/HGV RNA. For the Pygmies, GBV-C/HGV RNA-positive samples were found in 14 (8.0%) of the 174 from the Baka Pygmy in Cameroon and six (20.0%) of the 30 from the Aka Pygmy in Congo. There was no significant difference in the prevalence between any pairs of these three groups.

3.2. Detection of the GBV-C/HGV genotype

When the RFLP analysis was conducted for the 5′-UTR, the GBV-C/HGV genotype of the 20 isolates with GBV-C/HGV RNA from the Pygmies was found to be 18 for the GB type and two for the HG type. Of the 26 isolates positive for GBV-C/HGV RNA from the Bantu, all belonged to the

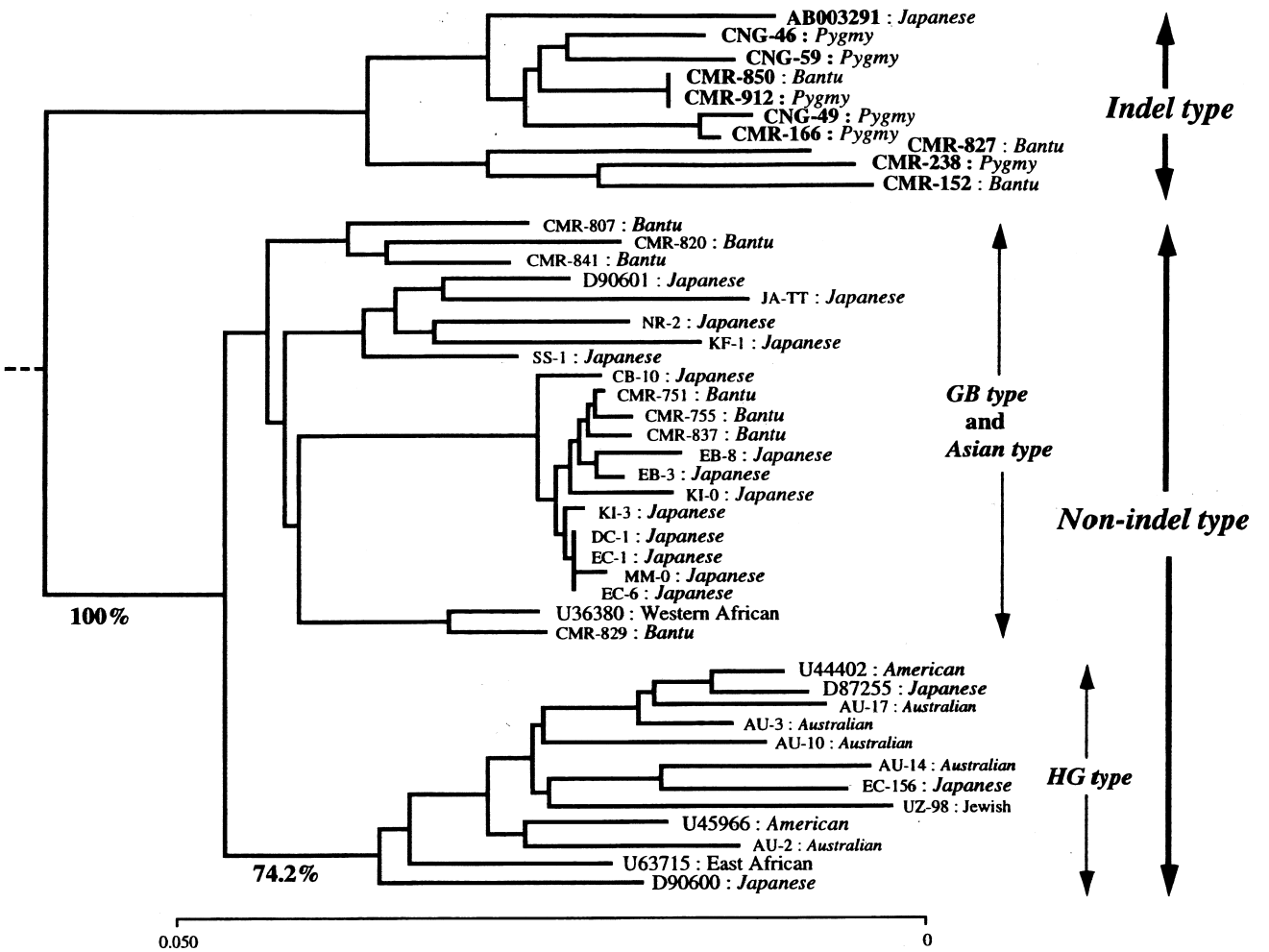

Fig. 2. Phylogenetic trees of the GBV-C/HGV NS5A region as constructed by the neighbor-joining method using all the isolates shown in Fig. 1. Isolates obtained in previous reports are indicated by their accession numbers from the DDBJ/EMBL/GenBank DNA databases. Isolates obtained from the Pygmies are indicated as 'Pygmy '. For the other isolates, the ethnic group are also shown. The horizontal bar indicates the number of amino acid substitutions per site. Ten isolates with the indel, shown in bold letters, fell within the same cluster with a bootstrap value of 100% as the Indel type.

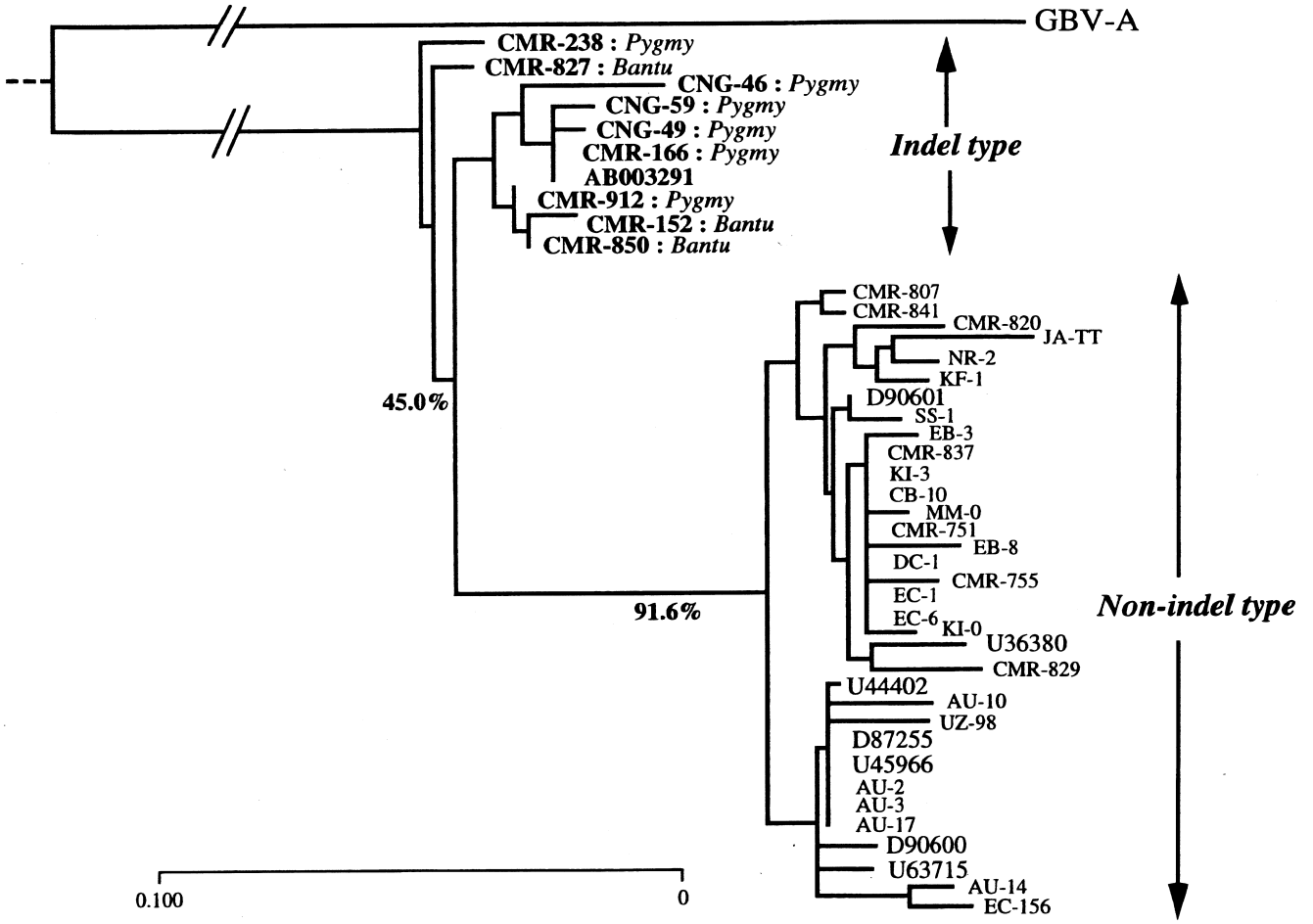

Fig. 3. Phylogenetic trees of the GBV-C/HGV NS5A region using GBV-A as an outgroup. Ten Indel types formed an outer group of the GBV-C/HGV isolates.

GB type. When non-African isolates were incorporated into the present analysis, a total of 94 isolates were classified into 46 GB types, 21 HG types, and 27 Asian types (Table 1).

3.3. Phylogenetic analysis of the NS5A region and African origin of GBV-C/HGV

Thirty-six isolates chosen randomly from 94 positive isolates (Table 1) were tested by RT-PCR using primers derived from the NS5A region. The 36 isolates consist of 16 GB types, 8 HG types, and 12 Asian types. When the amplicons were analyzed by electrophoresis on 3% agarose gels, there were two kinds of different bands, showing that there are the 'Indel' (i.e. insertion or deletion) type and 'non-Indel' type. To confirm the existence of these two types at the nucleotide sequence level, we determined nucleotide sequences for all 36 isolates. Of the 36 sequences, nine had a 36-nucleotide (corresponding to 12 amino acids (aa)) indel in the NS5A region (Fig. 1). Interestingly, all of these nine isolates were obtained from the Central African region and were identified as belonging to the GB type. This 12-aa indel was rich in serine residues, and just downstream of this indel there were similar sequences (Fig. 1). To examine if all the Indel types consist of the GB type only, all 94 positive isolates were tested by RT-PCR and analyzed by electrophoresis. As a result, 15 isolates were the Indel type and all of them were found to be the GB types that were mainly obtained from the African region.

The phylogenetic tree for the NS5A region indicated that there were two different clusters separated by a bootstrap value of 100%. It is of interest that one cluster consisted exclusively of Indel types, and the other cluster consisted of all the non-Indel types. More interestingly, the non-Indel types contained African and non-African isolates intermingled (Fig. 2). To determine a root of the phylogenetic tree, we used GBV-A as an outgroup. As shown in Fig. 3, the Indel type formed a distinct outer group of the GBV-C/HGV isolates.

Thus, we found that the 15 isolates with the 12-aa indel from the Pygmies and the Bantu, belonging to the Indel type, had apparently diverged earlier from the common ancestor than the non-Indel type. Along with the fact that the Pygmies and the Bantu diverged much earlier than non-Africans from the common ancestor, our results indicate that the GBV-C/HGV might be of African origin.

3.4. Phylogenetic analysis of NS3 region sequences

To elucidate the hypothesis that GBV-C/HGV was derived from Africa, we constructed a phylogenetic tree for the NS3 region, using GBV-C/HGV isolates consisting of 10 Indel types and 11 non-Indel types (namely, three GB types, five HG types, and three Asian types) and GBV-A as the outgroup. As a result, the 10 Indel types formed an outer group

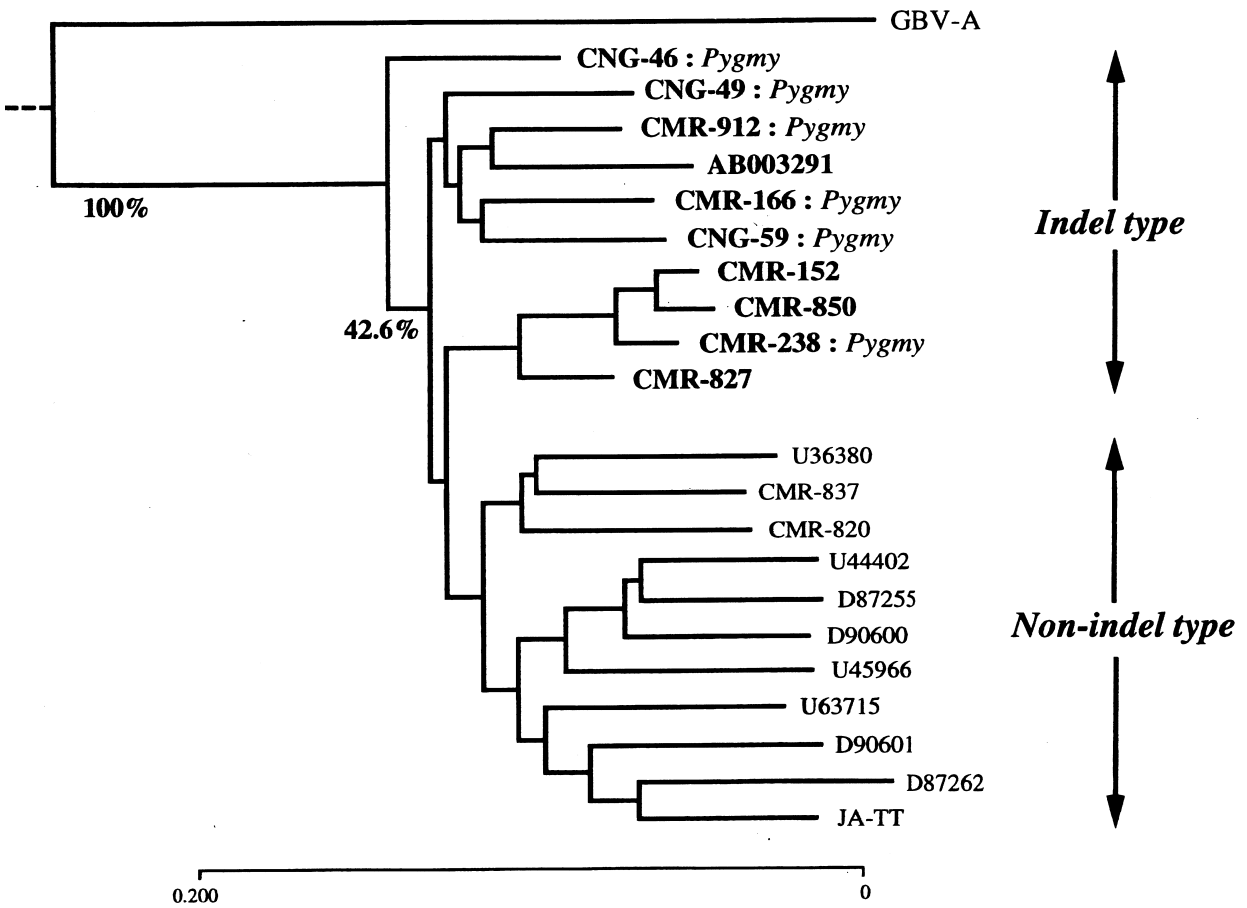

Fig. 4. Phylogenetic trees of the GBV-C/HGV NS3 region. We constructed a phylogenetic tree using 21 GBV-C/HGV isolates consisting of 10 Indel types (CMR-152, CMR-166, CMR-238, CMR-827, CMR-850, CMR-912, CNG-46, CNG-49, CNG-59 and AB003291) and 11 non-Indel types (three GB types: U36380, CMR-820 and CMR-837; five HG types: U44402, U45966, U63715, D87255 and D90600; three Asian types: D87262, D90601 and JA-TT) and GBV-A as an outgroup. The horizontal bar indicates the number of nucleotide substitutions per site.

of the GBV-C/HGV isolates (Fig. 4), which was consistent with the phylogenetic tree for the NS5A region. This observation reinforces our idea that GBV-C/HGV might be of African origin.

4. Discussion

Recently, GBV-C was isolated from West Africa for the first time [1]. It has been reported that GBV-C/HGV infection may not cause chronic liver disease [20,21]. Alternatively, GBV-C/HGV can infect human beings, but has little pathogenicity for humans at present. On the other hand, phylogenetic analyses using human mitochondrial DNA suggest that *Homo sapiens* has diverged from *Homo erectus* in Africa around two hundred thousand years ago and migrated out of Africa one hundred thousand years ago [22]. If GBV-C/HGV and humans have been able to coexist for a long time without pathogenicity, GBV-C/HGV might also have its origin in the African continent and have been carried out of Africa along with human migration. To elucidate the hypothesis that GBV-C/HGV was derived from Africa, molecular evolutionary analyses were performed using 94 GBV-C/HGV RNA-positive isolates from Central African regions and other regions.

In the Central African region, the prevalence of GBV-C/HGV was high, and most isolates belonged to the GB type. Of the 46 GBV-C/HGV RNA-positive isolates from the Central African region, 15 had a 12-aa indel in the NS5A region. When we constructed phylogenetic trees for the NS3, NS5A and NS5B regions (data not shown) using GBV-A as an outgroup, the Indel types from the Central African region formed an outer group of several GBV-C/HGV isolates. Moreover, 11 of the 15 Indel types were obtained from the Pygmies who are considered to be one of the oldest populations in Africa, and the prevalence of the Indel types in the Pygmies was significantly higher than that in the Bantu. Our data indicated that the Indel types from the Central African region, mainly from the Pygmies, might be the ancestors of GBV-C/HGV. For the isolates belonging to the non-Indel type, the 12-aa indel might have been deleted in the course of GBV-C/HGV evolution. This may be supported by the fact that deletions of amino acids have occurred about four times more frequently than insertions [23].

From our study, it would be reasonable to consider that GBV-C/HGV derives from Africa. A few questions still remain. First, the topology of the phylogenetic trees for the NS3 or NS5A regions was somewhat different from that for the 5′-UTR. It did not fully support our ideas. However, this may be explained possibly by the fact that these sequences were too short to construct phylogenetic trees or that recombinations among different genotypes might occur. The 12-aa deletion in the NS5A region of the GBV-C/HGV full sequen-

ces, however, was very characteristic and might have been one of the first significant events in the evolution of GBV-C/HGV. Second, we found that all the Indel types were obtained from the Central African region, but recently an isolate (AB003291) with a 12-aa indel in the NS5A region was detected in a Japanese patient [15]. When we reconstructed the phylogenetic trees for the 5′-UTR, NS3, NS5A and NS5B regions, the isolate certainly belonged to the cluster of the Indel type. Thus, it is possible that this Japanese isolate could have been brought from Africa to Japan in the past.

To elucidate the evolutionary origin of GBV-C/HGV, it is necessary to construct phylogenetic trees using more members of GBV-C/HGV isolates worldwide and to examine if the other non-A–G agents are more closely related to GBV-C/HGV than GBV-A. Moreover, if another flavivirus is discovered, it will give us useful information on the evolution of GBV-C/HGV.

References

[1] Simons, J.N., Leary, T.P., Dawson, G.J., Pilotmatias, T.J., Muerhoff, A.S., Schlauder, G.G., Desai, S.M. and Mushahwar, I.K. (1995) Nature Med. 1, 564–569.
[2] Linnen, J., Wages, J., Zhang-Keck, Z.-Y., Fry, K.E., Krawxzynski, K.Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J.W.-K., Young, L., Piatak, M., 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. and Kim, J.P. (1996) Science 271, 505–508.
[3] Mukaide, M., Mizokami, M., Orito, E., Ohba, K., Nakano, T., Ueda, R., Hikiji, K., Iino, S., Shapiro, S., Lahat, N., Park, Y.M., Kim, B.S., Oyunsuren, T., Rezieg, M., Al-Ahdal, M.N. and Lau, J.Y.N. (1997) FEBS Lett. 407, 51–58.
[4] Muerhoff, A.S., Simons, J.N., Leary, T.P., Erker, J.C., Chalmers, M.L., Pilot-Matias, T.J., Dawson, J., Desai, S.M. and Mushahwar, I.K. (1996) J. Hepatol. 25, 379–384.
[5] Muerhoff, A.S., Smith, D.B., Leary, T.P., Erker, J.C., Desai, S.M. and Mushahwar, I.K. (1997) J. Virol. 71, 6501–6508.
[6] Kondo, Y., Mizokami, M., Nakano, T., Kato, T., Ohba, K., Orito, E., Ueda, R., Mukaide, M., Hikiji, K., Oyunsuren, T. and Cooksley, W.G. (1997) Virus Res. (in press).
[7] Kato, T., Mizokami, M., Nakano, T., Kondo, Y., Ohba, K., Orito, E., Ueda, R., Mukaide, M., Gurtsevitch, V., Syrtsev, A., Ruzibakiev, R., Abdurakhnamov, M., Yamashita, M. and Hayami, M. (1997) Virus Res. 48, 81–87.
[8] Wu, R.-R., Mizokami, M., Cao, K., Nakano, T., Ge, X.-M., Wang, S.-S., Orito, E., Ohba, K., Mukaide, M., Hikiji, K., Lau, J.Y.N. and Iino, S. (1996) J. Infect. Dis. 175, 168–171.
[9] Ohba, K., Mizokami, M., Lau, J.Y.N., Orito, E., Ikeo, K. and Gojobori, T. (1996) FEBS Lett. 378, 232–234.
[10] Muerhoff, A.S., Leary, T.P., Simons, J.N., Pilot-Matias, T.J., Dawson, G.J., Erker, J.C., Chalmers, M.L., Schlauder, G.G., Desai, S.M. and Mushauhwar, I.K. (1995) J. Virol. 69, 5621–5630.
[11] 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. and Mushahwar, I.K. (1995) J. Med. Virol. 48, 60–67.
[12] Erker, J.C., Simons, J.N., Muerhoff, A.S., Leary, T.P., Chalmers, M.L., Desai, S.M. and Mushahwar, I.K. (1996) J. Gen. Virol. 77, 2713–2720.
[13] Nakao, H., Okamoto, H., Fukuda, M., Tsuda, F., Mitsui, T., Masuko, K., Iizuka, H., Miyakawa, Y. and Mayumi, M. (1997) Virology 233, 43–50.
[14] Okamoto, H., Nakao, H., Inoue, T., Fukuda, M., Kishimoto, J., Lizuka, H., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1997) J. Gen. Virol. 78, 737–745.
[15] Takahashi, K., Hijikata, M., Hino, K. and Mishiro, S. (1997) Hepatol. Res. 8, 139–148.
[16] Ina, Y. (1994) CABIOS 10, 11–12.
[17] Gojobori, T., Ishii, K. and Nei, M. (1982) J. Mol. Evol. 18, 414–423.
[18] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[19] Felsentein, J. (1985) Evolution 39, 783–791.
[20] Kao, J.H., Chen, P.J., Lai, M.Y., Chen, W., Liu, D.P., Wang, J.T., Shen, M.C. and Chen, D.S. (1997) Gastroenterology 112, 1265–1270.
[21] Nakatsuji, Y., Shih, J.W.-K., Tanaka, E., Kiyosawa, K., Wages Jr., J., Kim, J.P. and Alter, H.J. (1996) J. Viral Hepatitis 3, 307–316.
[22] Wilson, A.C. and Cann, R.L. (1992) Sci. Am. 266, 68–73.
[23] De Jong, W.W. and Ryden, L. (1981) Nature 290, 157–159.