Virology **277,** 330–335 (2000)
doi:10.1006/viro.2000.0621, available online at http://www.idealibrary.com on IDEAL®

Full-Length Nucleotide Sequence of a Simian TT Virus Isolate Obtained from a Chimpanzee: Evidence for a New TT Virus-like Species

Tomoko Inami,*,† Takashi Obara,‡ Mitsuhiko Moriyama,† Yasuyuki Arakawa,† and Kenji Abe*,[1]

**Department of Pathology, National Institute of Infectious Diseases, Tokyo; ‡Diagnostic Division, Eisai Co., Ltd., Ibaraki; and †3rd Department of Internal Medicine, Nihon University School of Medicine, Tokyo, Japan*

Received May 18, 2000; returned to author for revision August 2, 2000; accepted August 29, 2000

Recently, we identified TT virus (TTV) isolates from nonhuman primates and named them simian TTV (s-TTV). To characterize the genomic structure of these isolates in more detail, the full-length nucleotide sequence of the s-TTV isolate (designated s-TTV CH65-1), recovered from a chimpanzee born in West Africa, was amplified by nested PCR with inverted primers deduced from the untranslated region of s-TTV DNA. CH65-1 was composed of 3899 nucleotides (nt) and had two open reading frames (ORF) spanning 2295 nt (ORF1) and 402 nt (ORF2). The sequence had only 52.3% similarity to the prototype TA278 human isolate. Phylogenetic analysis demonstrated that CH65-1 was distinct from the human TTV isolates. These results suggested that s-TTV may represent a new TTV-like viral species or genus.

INTRODUCTION

In 1997, the genome of a novel DNA virus, named the TT virus (TTV), was discovered in a patient with acute posttransfusion hepatitis by representational difference analysis (Nishizawa *et al.,* 1997; Okamoto *et al.,* 1998). TTV is an unenveloped, circular, single-stranded DNA virus (having 3852 nucleotides in the full-length sequence), with an isopycnic density of 1.31–1.34 g/ml in CsCl (Miyata *et al.,* 1999; Mushahwar *et al.,* 1999). The TTV genome has two or three possible open reading frames (ORF), capable of encoding 770 (ORF1), 202 (ORF2), and 105 (ORF3) amino acids (Miyata *et al.,* 1999). The genome structure and its banding in buoyant density gradient centrifugation suggest that TTV might be most related to the *Circoviridae* virus from among the known animal virus families (Miyata *et al.,* 1999; Mushahwar *et al.,* 1999; Takahashi *et al.,* 1998a). Despite being a DNA virus, the TTV has a wide range of sequence divergence, allowing classification into several genotypes (Abe *et al.,* 1999; Mushahwar *et al.,* 1999; Takahashi *et al.,* 1998a; Tanaka *et al.,* 1998). TTV sequences can be detected in sera and liver tissues from liver disease patients, suggesting that TTV would be responsible for some of the acute and chronic liver disease cases of unknown etiology (Charlton *et al.,* 1998; Okamoto *et al.,* 1998). On the other hand, it has been reported that TTV infection does not induce significant liver damage (Naoumov *et al.,* 1998). We have reported a very high prevalence of TTV in general populations worldwide, suggesting that this virus may be a common DNA virus with no clear disease association in humans (Abe *et al.,* 1999). However, the epidemiology, clinical significance, and transmission patterns of TTV remain unclear.

Chimpanzees are susceptible to infection with TTV (Mushahwar *et al.,* 1999). Furthermore, TTV DNA sequences have been found in nonhuman primates and farm animals (Leary *et al.,* 1999; Okamoto *et al.,* 2000; Romeo *et al.,* 2000; Verschoor *et al.,* 1999). Very recently, we identified TTV isolates from nonhuman primates, including chimpanzees and crab-eating macaques, and tentatively named them simian TTV (s-TTV) (Abe *et al.,* 2000). Interestingly, our results revealed that the TTV isolates obtained from simians were distinct from the human TTV isolates. In order to characterize the s-TTV genome in detail, we carried out the complete genome sequencing of this isolate.

RESULTS

The sequence of the CH65-1 isolate was determined by combining three fragments (Fig. 1). The sequences of primers used for this study are shown in Table 1. We cloned the full-length nucleotide sequence of the s-TTV isolate (designated s-TTV CH65-1) recovered from a chimpanzee. The CH65-1 isolate was composed of 3899 nucleotides (nt) and had two putative open reading frames (ORF; >100 nt) spanning 2295 nt (ORF1) and 402

The nucleotide sequence data reported in this paper have been submitted to the DDBJ, EMBL, and GenBank nucleotide sequence databases with the following accession numbers: AB037926 (complete genome for the s-TTV CH65-1 isolate) and AB035154 through AB035171 (partial sequence in 5′ untranslated region for the other s-TTV isolates).

[1] To whom reprint requests should be addressed at the Department of Pathology, National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku-ku, Tokyo 162-8640, Japan. Fax: (81) 3-5285-1189. E-mail: kenjiabe@nih.go.jp.

0042-6822/00 $35.00

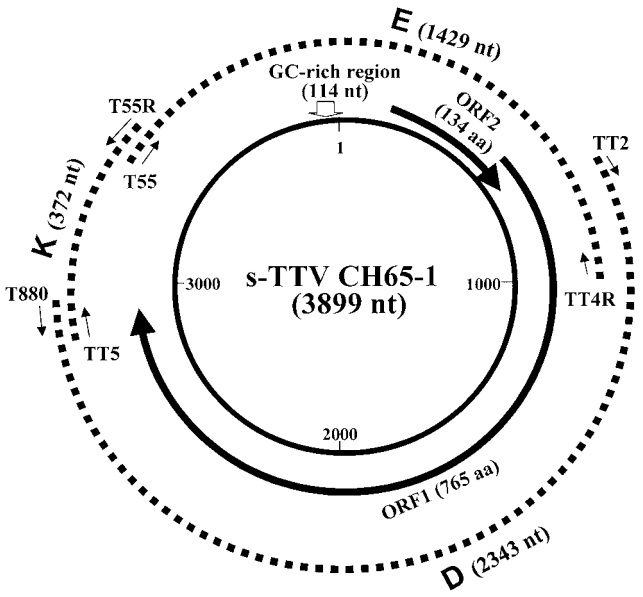

FIG. 1. Strategy for the full genome sequencing of the s-TTV CH65-1 isolate. The sequence of the CH65-1 isolate was determined by combining three fragments. Fragment E was amplified by the inverted PCR. The primer sequences are as follows: (1) for fragment D, 5′-GTC AAG GGG CAA TTC GGG CTC-3′ (T855, sense primer, design from TA278, nt 204–224) and 5′-GCT K(G or T)CG CTC GGA GTG CTT AG-3′ (T881, antisense primer, designed from TA278, nt 3126–3145) for the outer primer pairs, and 5′-AGA CGC AGA CCT GCT AGA CG-3′ (TT2, sense primer, designed from TA278, nt 673–692) and 5′-GAC AAG TGA AAS(G or C) TCC CAC GG-3′ (T880, antisense primer, designed from TA278, nt 3085–3104) for the inner primer pairs. (2) For fragment E, 5′-CGA AAG TGA GTG GGG CCA GAC-3′ (T55, sense primer, designed from TA278, nt 3337–3357) and 5′-CGC ACC ACA GGA TGG GAA AC-3′ (TT4R, antisense primer, designed from CH65-1, nt 901–920). (3) For fragment K, 5′-GTT CGG CTC ACC ACT AAC TG-3′ (TT4, sense primer, designed from CH65-1, nt 1396–1414) and 5′-GTC TGG CCC CAC TCA CTT TCG-3′ (T55R, antisense primer, designed from TA278, nt 3357–3337) for the outer primer pairs, and 5′-ATC GCC CCG TCC GCT CTT TC-3′ (TT5, sense primer, designed from CH65-1, nt 3040–3059) and 5′-GTC TGG CCC CAC TCA CTT TCG-3′ (T55R, antisense primer, designed from TA278, nt 3357–3337) for the inner primer pairs. aa, amino acids.

nt (ORF2). Fragment E was obtained by an inverted PCR. This result indicated that CH65-1 had a circular genome as does TTV in humans. ORF1 and ORF2 overlapped for 142 nt (nt 623–765) in CH65-1. ORF1 in CH65-1 was shorter by 5 amino acids and showed only 35% identity to the prototype TA278 human isolate at the amino acid level. In addition, there was a 114-nt GC-rich region (nt 3788-3; GC content = 90%), as seen in the human TTV genome. TATA boxes such as ATATAA and TATATA were found at nt positions 85–90. Hydrophilicity profiles of both ORFs between CH65-1 and TA278 were compared by the method of Hopp and Woods (Hopp and Woods, 1981). Although there were low similarities in the amino acid sequences of 35% for ORF1 and 43% for ORF2 (Table 1), both isolates had very similar hydrophilicity patterns (Fig. 2). When compared to five other reported TTV isolates with full-length genome sequences, CH65-1 showed overall identities of only 48 to 54% at the nucleotide level, thereby indicating that this isolate was a different virus (Table 1). The sequence similarity was much higher, 71%, within the untranslated region (UTR). The alignment of UTR having a GC-rich region is shown in Fig. 3. The results indicated that a ATF/CREB site, a AP2 site, a SP1 site, and a TATA box were well conserved when compared with reported TTV isolates in humans. The GC-rich region in this region showed the highest similarity at 89% (Table 1). In contrast, the similarity of two coding regions was much lower, being 49 to 51% in ORF1 and 48 to 65% in ORF2 at the nucleotide level. Furthermore, CH65-1 did not show any significant homology with reported sequences. Phylogenetic analysis demonstrated that the CH65-1 isolate was clearly distinct from the TTV found in humans, and this difference was strongly supported by bootstrap analysis (Fig. 4). We have reported previously that s-TTV can be divided into genetically distinct genotypes by phylogenetic analysis of the 5′-UTR sequence (Abe *et al.,* 2000). By the same analysis of the short sequence of 5′-UTR, CH65-1 was located in the middle, between genotypes 1 and 2 (Fig. 5). Interestingly, in this region, CH65-1 had characterization of the nucleotide sequences in both type 1 and type 2 s-TTV. This suggests that CH65-1 appeared to be a

TABLE 1

Percentages of Nucleotide and Amino Acid Identities within Full Genome and two ORFs of s-TTV CH65-1 for Five Reported TTVs in Humans and CAV

Isolate	Accession No.	Length (base)	Full	ORF1 nt	ORF1 aa	ORF2 nt	ORF2 aa	UTR (GC-rich region)
TA278	AB017610	3852	52	49	35	48	43	70 (89)
GH1	AF122913	3852	53	50	34	51	43	71 (88)
JA1	AF122916	3839	54	49	41	52	43	71 (87)
SANBAN	AB025946	3808	52	51	51	65	53	62 (68)
TUS01	AB017613	3818	48	49	41	57	44	68 (79)
CAV[a]	D31965	2319	44					

[a] Chicken anemia virus (Kato *et al.,* 1995).

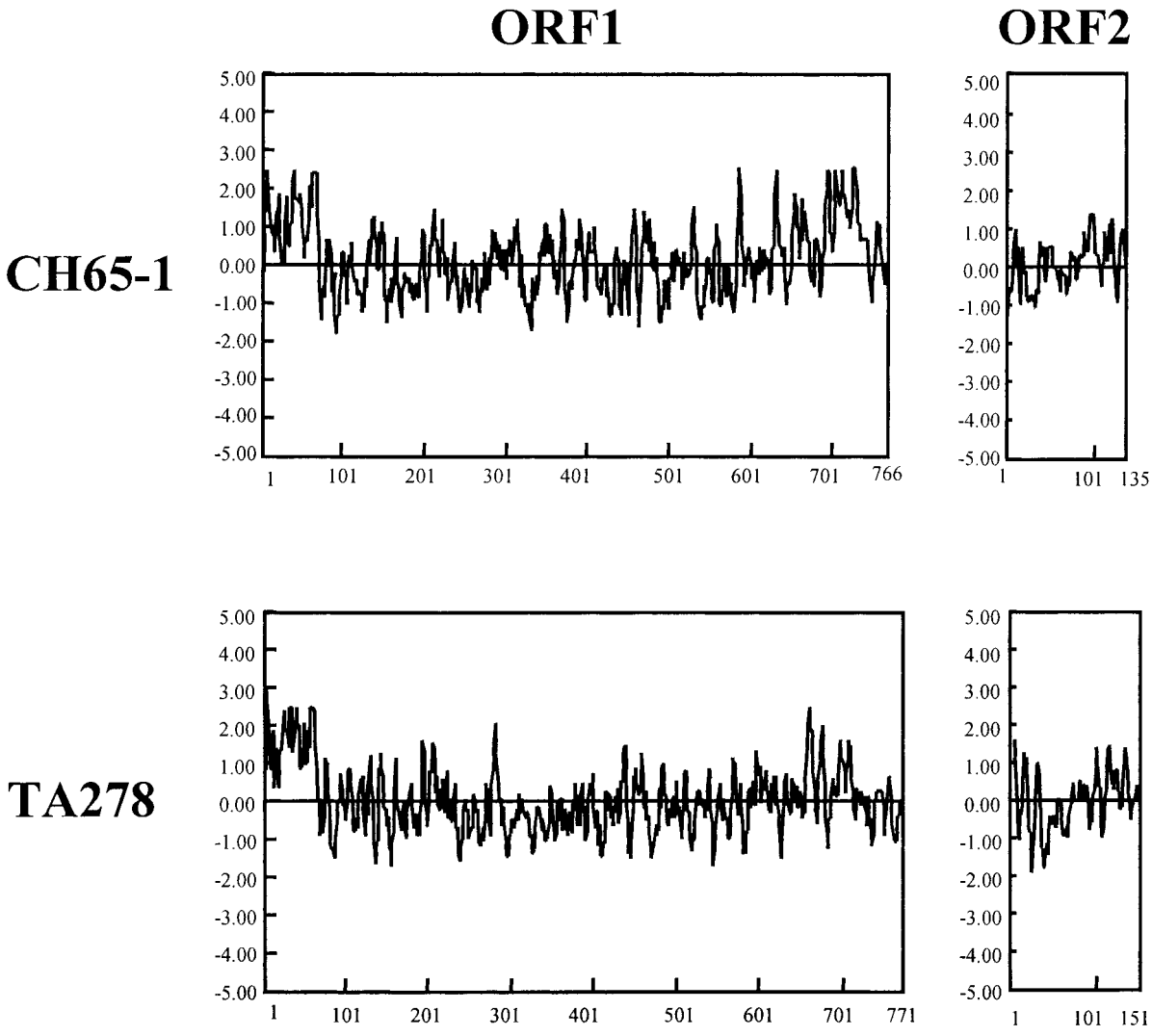

FIG. 2. Comparison of hydrophilicity profiles of proteins encoded by the two ORFs of s-TTV and TTV. A hydrophilicity score of six amino acids was calculated by the method of Hopp and Woods (1981) for ORF1 and ORF2 in both isolates.

recombinant between the type 1 and type 2 strains of s-TTV.

DISCUSSION

TTV viremia is widespread, with a very high incidence in general populations worldwide (Abe *et al.,* 1999). This suggests that TTV is a common virus and may be a nonpathogenic DNA virus in humans. Recently, TTV DNA sequences have been found in nonhuman primates and farm animals (Leary *et al.,* 1999; Okamoto *et al.,* 2000; Romeo *et al.,* 2000; Verschoor *et al.,* 1999). Very recently, we reported a high prevalence of s-TTV infection in captured chimpanzees and crab-eating macaques. These findings suggest that s-TTV is widespread among wild chimpanzees living in West Africa. All the animals infected with s-TTV were clinically healthy. Thus s-TTV may be nonpathogenic in chimpanzees and crab-eating monkeys as well as in TTV-infected human individuals, although the pathogenic role of this virus still remains to be investigated. It is known that the TTV genome has a markedly wide range of sequence divergence, in which it is classified into at least 16 genotypes separated by a difference of >30% within a partial sequence in the coding region (N22 region) (Okamoto *et al.,* 1999a). In the present study, the entire nucleotide sequence was determined for the s-TTV CH65-1 isolate recovered from a wild chimpanzee in West Africa. It had a circular genome structure similar to that of the TTV in humans and consisted of 3899 nt. From the degree of singularity of CH65-1 compared to the prototype TTV isolate and known reported TTVs, we were able to confirm that CH65-1 should belong to a new viral species or subspecies of TTV. The genetic distance between CH65-1 and TTV in humans is too great to allow them to be considered as different genotypes. This argument may

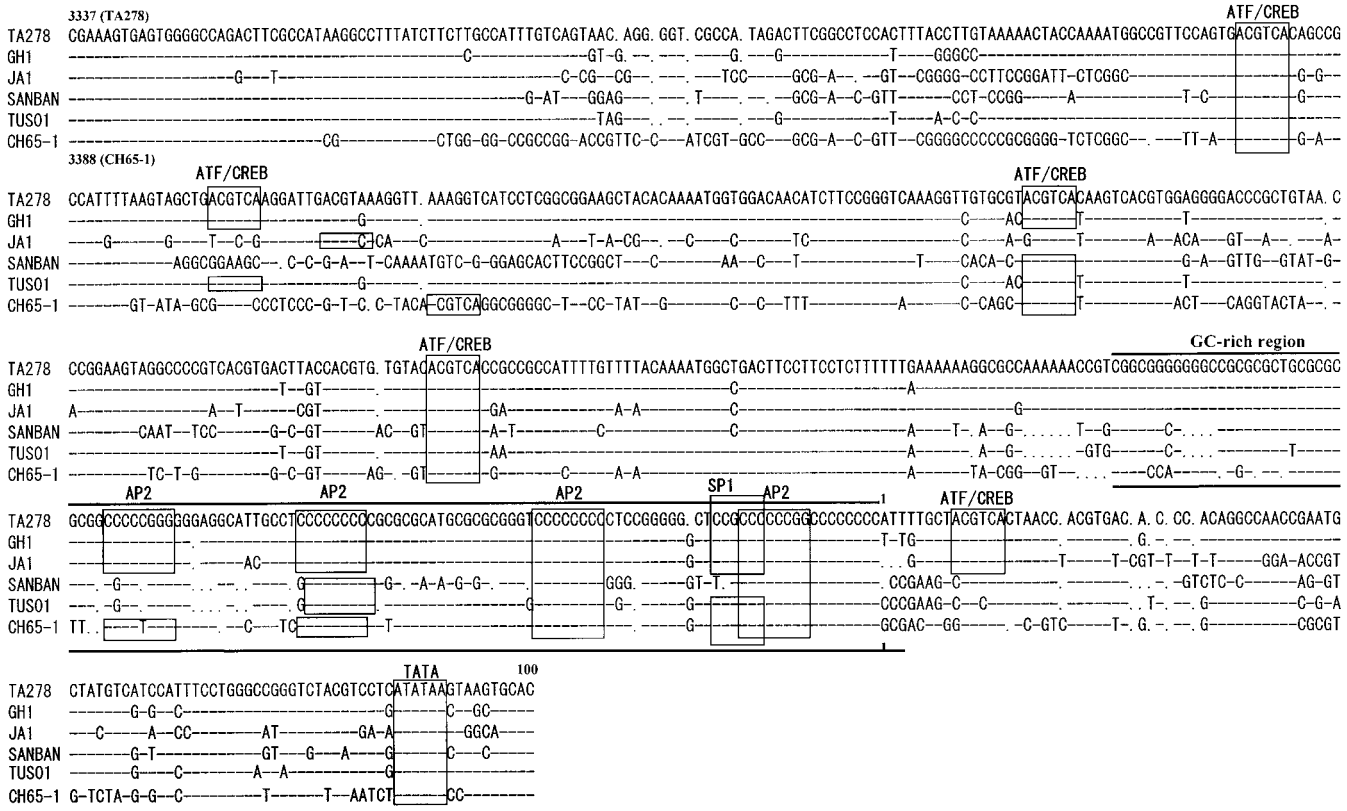

FIG. 3. Alignment of nucleotide sequences of TTV and s-TTV CH65-1 of 3′-UTR. The sequence of TA278 is indicated at the top for comparison. Dashes indicate the same nucleotides as those in TA278, and dots represent deletions. The GC-rich region is indicated with thick lines. The consensus sequences of "ACGTCA" for "ATF/CREB," "CCC[AC]N[CG][CG][CG]" for "AP2," "GGGCGG/CCGCCC" for "SP1," and "ATATAA, TATATA" for "TATA" are boxed.

further be supported by the fact that the amino acid sequence identity between CH65-1 and TTV was less than 53% in both ORF1 and ORF2, although the nucleotide sequence identity between SANBAN and CH65-1 showed a 65% similarity. These findings were also supported by phylogenetic analysis based on the full-length nucleotide sequence data. In any case, to determine the relationship between s-TTV and human TTV further, more isolates of each group will need to be sequenced over the entire genome. Studies of the TTV family are important in determining the origin of the virus and may provide useful information and models for studies of the pathogenicity and diversity of this virus. In fact, Gao *et al.* reported that a lineage of simian immunodeficiency virus infecting chimpanzees was closely related to all groups of human immunodeficiency virus type 1 and speculated that chimpanzees are the primary reservoir of this virus (1999).

In conclusion, we have cloned the entire nucleotide sequence of s-TTV isolate recovered from a chimpanzee. Based on analysis of full-length sequence data, we confirmed that s-TTV may represent a new TTV-like viral species or genus, although it is closely related to human TTV.

MATERIALS AND METHODS

Serum sample

We used serum samples obtained from a chimpanzee that was seropositive for s-TTV DNA by PCR as reported previously (Abe *et al.,* 2000). The serum sample was negative for TTV DNA by PCR with primers deduced from the N22 region in ORF1 of TA278, but, nevertheless, was positive by a PCR set using a primer deduced from a 5′-UTR (Takahashi *et al.,* 1998b). The chimpanzee used for this study was born in West Africa and imported into Japan in 1979. We used serum samples that were obtained from the animal immediately after its arrival in Japan at quarantine. The age of the animal was unknown. The chimpanzee had not been inoculated previously with human serum, any hepatitis viruses, or other serum or blood products. The serum samples were kept at −40°C or below until tested.

Strategy for full-length nucleotide sequencing

DNA was extracted from 100 μl of serum sample with a nucleic acid extraction kit (SepaGene RV-R, Sanko Junyaku Co., Ltd.) as directed by the manufacturer. In brief, to obtain the full-length sequence, the CH65-1

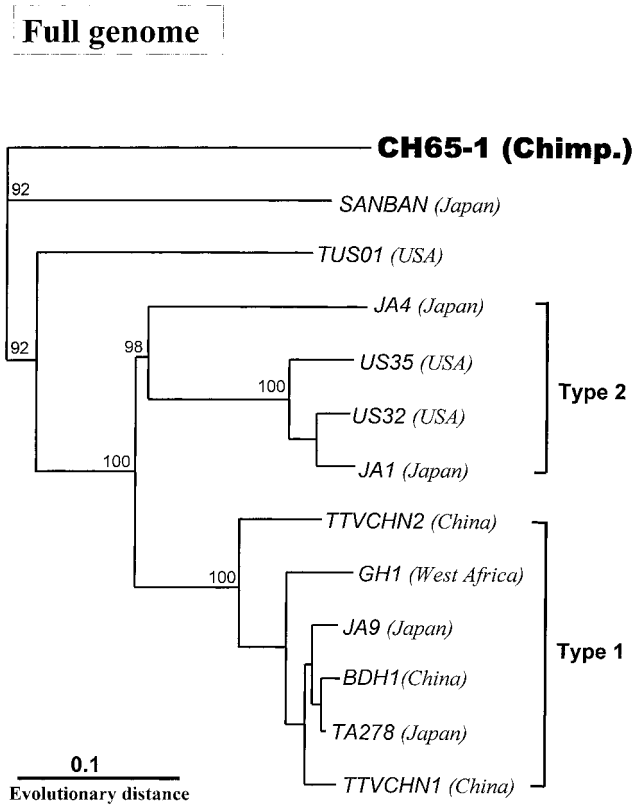

FIG. 4. A phylogram generated by neighbor-joining analysis of genetic distances in the full-length sequence of TTV and s-TTV. The numbers at junctures are percentages of bootstrap replicates supporting these branches. Isolates from the database are shown in italics.

isolate was amplified by nested PCR consisting of three fragments (Fig. 1). At first, the sequence of fragment D spanning 2343 nt in CH65-1 was determined by nested PCR with LA Taq (TaKaRa) and primers deduced from TA278, while those of fragments E (1429 nt) and K (372 nt) were determined by a single round of PCR with LA Taq using GC buffer (TaKaRa) and seminested PCR with LA Taq (TaKaRa), respectively. As shown in Fig. 1, two primers for fragments E and K were deduced from fragment D of CH65-1 (TT5 and TT4R) and the others from TA278. Fragment E was obtained by an inverted PCR. The complete sequence of the circular genome of CH65-1 was determined by combining these three overlapping sequences.

Cloning and nucleotide sequencing

PCR products were separated by 1% agarose gel electrophoresis and purified using the QIAquick gel extraction kit (Qiagen Inc.). Recovered PCR products were subcloned using a pBluescript II SK(−) vector (Stratagene) through the Eco RV site. Nucleotide sequencing was performed using the ABI PRISM BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Perkin–Elmer). Sequences of amplified cDNA were determined using a sequencer (ABI PRISM™ 310 Genetic Analyzer; Applied Biosystems). Alternatively, the GC-rich region in fragment E was also determined after the method of Maxam and Gilbert (1980).

Phylogenetic analysis

Nucleotide sequences were multiple aligned using CLUSTAL W version 1.4. The distance matrix of the nucleotide substitutions within each clone was estimated by the eight-parameter method (Rzhetsky, 1995), and phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei, 1987) from the matrix. These procedures were computed using Phylo_win version 1.2 (Galtier *et al.,* 1996) on a DEC alpha 2000 server, and the trees were drawn by TreeView version 1.5 (Page, 1996). The reliability and topology of each tree branch was tested by bootstrap analysis (Billis and Bull, 1993) of the data of 1000 bootstrap resamplings of the columns in the full-genome sequence alignment. Sequences of TTV isolates obtained from databases were used to compare the sequences of the isolates in the present study. The isolates' names and accession numbers and references of the reported sequences were as follows: TA278(AB017610) (Miyata *et al.,* 1999; Okamoto *et al.,* 1998), GH1(AF122913) (Mushahwar *et al.,* 1999),

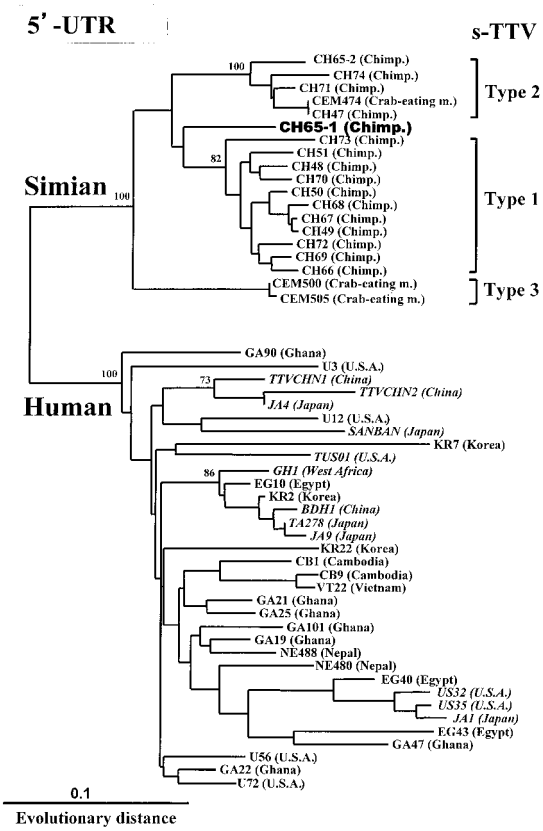

FIG. 5. A phylogram generated by neighbor-joining analysis of genetic distances in the 5′-UTR sequences of TTV and s-TTV.

SANBAN(AB025946) (Hijikata *et al.,* 1999), TUS01-(AB017613) (Okamoto *et al.,* 1999b), JA1(AF122916), JA4(AF122917), JA9(AF122915), US32 (AF122921), US35-(AF122920) (Erker *et al.,* 1999), TTVCHN1 (AF079173), TTVCHN2 (AF129887), and BDH1 (AF116842) for human TTV isolates.

ACKNOWLEDGMENTS

We thank Dr. Tetsutaro Sata for his continuous encouragement during this study and Drs. Hideo Naito and Naoto Aiba, National Institute of Infectious Diseases, for their kind help.

REFERENCES

Abe, K., Inami, T., Asano, K., Miyoshi, C., Masaki, N., Hayashi, S., Ishikawa, K., Takebe, Y., Win, K. M., Zayadi, A. R., Han, K. H., and Zhang, D. Y. (1999). TT virus infection is widespread in the general populations from different geographic regions. *J. Clin. Microbiol.* **37,** 2703–2705.

Abe, K., Inami, T., Ishikawa, K., Nakamura, S., and Goto, S. (2000). TT virus infection in nonhuman primates and characterization of viral genome: Identification of simian TT virus isolates. *J. Virol.* **74,** 1549–1553.

Billis, D. M., and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42,** 182–192.

Charlton, M., Adjei, P., Poterucha, J., Zein, N., Moore, B., Therneau, T., Krom, R., and Wiesner, R. (1998). TT-virus infection in north American blood donors, patents with fulminant hepatic failure, and cryptogenic cirrhosis. *Hepatology* **28,** 839–842.

Erker, J. C., Leary, T. P., Desai, S. M., Chalmers, M. L., and Mushahwar, I. K. (1999). Analysis of TT virus full-length genomic sequence. *J. Gen. Virol.* **80,** 1743–1750.

Galtier, N., Gouy, M., and Gautier, C. (1996). SeaView and Phylo_win, two graphic tools for sequence. *Comput. Appl. Biosci.* **12,** 543–548.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes. Nature* **397,** 436–441.

Hijikata, M., Takahashi, K., and Mishiro, S. (1999). Complete circular DNA genome of a TT virus variant (isolate name SANBAN) and 44 partial ORF2 sequence implicating a great degree of diversity beyond genotypes. *Virology* **260,** 17–22.

Hopp, T. P., and Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78,** 3824–3828.

Kato, A., Fujino, M., Nakamura, T., Ishihama, A., and Otaki, Y. (1995). Gene organization of chicken anemia virus. *Virology* **209,** 480–488.

Leary, T. P., Erker, J. C., Chalmers, M. L., Desai, S. M., and Mushahwar, I. K. (1999). Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals. *J. Gen. Virol.* **80,** 2115–2120.

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65,** 499–560.

Miyata, H., Tsunoda, H., Kazi, A., Yamada, A., Khan, M. A., Murakami, J., Kamahora, T., Shiraki, K., and Hino, S. (1999). Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. *J. Virol.* **73,** 3582–3586.

Mushahwar, I. K., Erker, J. C., Muerhoff, A. S., Leary, T. P., Simons, J. N., Birkenmeyer, L. G., Chalmers, M. L., Pilot-Matias, T. J., and Dexai, S. M. (1999). Molecular and biophysical characterization of TT virus: Evidence for a new virus family infecting humans. *Proc. Natl. Acad. Sci. USA* **96,** 3177–3182.

Naoumov, N. V., Petrova, E. P., Thomas, M. G., and Williams, R. (1998). Presence of a new described human DNA virus (TTV) in patients with liver disease. *Lancet* **352,** 195–197.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y., and Mayumi, M. (1997). A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem. Biophys. Res. Commun.* **241,** 92–97.

Okamoto, H., Fukuda, M., Tawara, A., Nishizawa, T., Itoh, Y., Hayasaka, I., Tsuda, F., Tanaka, T., Miyakawa, Y., and Mayumi, M. (2000). Species-specific TT viruses and cross-species infection in nonhuman primates. *J. Virol.* **74,** 1132–1139.

Okamoto, H., Nishizawa, T., Kato, N., Ukita, M., Ikeda, H., Iizuka, H., Miyakawa, Y., and Mayumi, M. (1998). Molecular cloning and characterization of a novel DNA virus (TTV) associated with posttransfusion hepatitis of unknown etiology. *Hepatol. Res.* **10,** 1–16.

Okamoto, H., Nishizawa, T., Ukita, M., Takahashi, M., Fukuda, M., Iizuka, H., Miyakawa, Y., and Mayumi, M. (1999b). The entire nucleotide sequence of a TT virus isolate from the United States (TUS01): Comparison with reported isolates and phylogenetic analysis. *Virology* **259,** 437–448.

Okamoto, H., Takahashi, M., Nishizawa, T., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y., and Mayumi, M. (1999a). Marked genomic heterogeneity and frequent mixed infection of TT virus demonstrated by PCR with primers from coding and noncoding regions. *Virology* **259,** 428–436.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12,** 357–358.

Romeo, R., Hegerich, P., Emerson, S. U., Colombo, M., Purcell, R. H., and Bukh, J. (2000). High prevalence of TT virus (TTV) in naive chimpanzees and in hepatitis C virus-infected humans: Frequent mixed infections and identification of new TTV genotypes in chimpanzees. *J. Gen. Virol.* **81,** 1001–1007.

Rzhetsky, A. N. (1995). Tests of applicability of several substitution models for DNA sequence data. *Mol. Biol. Evol.* **12,** 131–151.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–423.

Takahashi, K., Hoshino, H., Ohta, Y., Yoshida, N., and Mishiro, S. (1998b). Very high prevalence of TT virus (TTV) infection in general population of Japan revealed by a new set of PCR primers. *Hepatol. Res.* **12,** 233–239.

Takahashi, K., Ohta, Y., and Mishiro, S. (1998a). Partial ~2.4kb sequence of TT virus genome from eight Japanese isolates: Diagnostic and phylogenetic implications. *Hepatol. Res.* **12,** 111–120.

Tanaka, Y., Mizokami, M., Orito, E., Ohno, T., Nakano, T., Kato, T., Kato, H., Mukaide, M., Park, Y. M., Kim, B. S., and Ueda, R. (1998). New genotypes of TT virus (TTV) and a genotyping assay based on restriction fragment length polymorphism. *FEBS Lett.* **437,** 201–206.

Verschoor, E. J., Langenhuijzen, S., and Heeney, J. L. (1999). TT viruses (TTV) of non-human primates and their relationship of the human TTV genotypes. *J. Gen. Virol.* **80,** 2491–2499.