Identification and genomic characterization of a novel human torque teno virus of 3.2 kb

Masashi Ninomiya,1,2 Tsutomu Nishizawa,1,3 Masaharu Takahashi,1 Felipe R. Lorenzo,1 Tooru Shimosegawa2 and Hiroaki Okamoto1

1Division of Virology, Department of Infection and Immunity, Jichi Medical University School of Medicine, Tochigi-Ken 329-0498, Japan
2Department of Gastroenterology, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan
3International Research and Educational Institute for Integrated Medical Sciences, Tokyo Women’s Medical University, Tokyo 162-8666, Japan

In 1997, a novel DNA virus, unrelated to the known human viruses, was isolated by representational difference analysis from the serum of a patient with post-transfusion hepatitis of unknown aetiology, and it was tentatively named as TT virus (TTV) after the initials (T. T.) of the index patient (Nishizawa et al., 1997; Okamoto et al., 1998). After the discovery of the initial TTV isolate, many TTV variants with marked genetic variability were identified and they segregated into at least 39 genotypes or five major genetic groups (Erker et al., 1999; Okamoto et al., 1999a, 2004; Hallett et al., 2000; Khudyakov et al., 2000; Takahashi et al., 2000a; Peng et al., 2002). In 2000, a small virus that was distantly related to TTV and provisionally named as TTV-like mini virus was discovered by using PCR with TTV-specific primers that partially matched homologous sequences in TTV-like mini virus (Takahashi et al., 2000b). Recently, the International Committee on Taxonomy of Viruses officially designated TTV and TTV-like mini virus as torque teno virus (TTV) and torque teno mini virus (TTMV), respectively, and classified them into a novel floating genus, Anellovirus (Biagini et al., 2005). TTV and TTMV are both unenveloped, small spherical particles with a circular single-stranded DNA genome of 3.8–3.9 and 2.8–2.9 kb, respectively. They share a similar genomic organization with four open reading frames (ORF1–ORF4) and a region of 80–160 nt with high G+C content (approx. 90 mol%), which have a high degree of similarity among the extensively divergent TTV or TTMV variants (Miyata et al., 1999; Mushahwar et al., 1999; Okamoto et al., 1999a, 2004; Takahashi et al., 2000b; Bendinelli et al., 2001; Biagini et al., 2001; Hino, 2002). Anellovirus strains have also been detected in non-human primates (chimpanzees, macaques, tamarins and douroucoulis), tupaia, cats, dogs and farm animals (Leary et al., 1999; Verschoor et al., 1999; Cong et al., 2000; Inami et al., 2000; Okamoto et al., 2000a, b, 2001b, 2002; Thom et al., 2003). Recently, two new viruses named small anellovirus 1 (SAV1) and small anellovirus 2 (SAV2) were isolated from the sera of patients with acute viral infection syndrome in the USA using DNase sequence-independent single-primer amplification (Jones et al., 2005). SAV1 possessed genomic DNA of 2249 nt with three putative ORFs, while SAV2 had genomic DNA of 2635 nt with five ORFs. Although the number of ORFs differed, these two viruses (collectively, SAVs) were provisionally classified as anelloviruses on the basis of the circular nature of the genomic DNA, and the presence of regions homologous to TTV and TTMV in the largest ORF (ORF1) and non-coding region. Although
it was reported that SAV DNA was detectable in sera from patients with hepatitis C and/or apparently healthy blood donors in Italy and France (Andreoli et al., 2006; Biagini et al., 2006), the geographical distribution of SAVs and their genomic variability remain unclear.

To investigate the presence of SAVs in Japan, specific primers amplifying a 925 nt SAV1 sequence and other primers amplifying a 1129 nt SAV2 sequence were designed. Serum samples obtained from 218 Japanese patients with haemophilia who were infected with blood-borne viruses including hepatitis B virus (4.6 %), hepatitis C virus (83.9 %), human immunodeficiency virus type 1 (35.3 %) and/or TTV (100 %) were subjected to the two PCR assays for the detection of SAV1 and SAV2 DNAs. To amplify the 925 nt SAV1 sequence, the primers NG696 (sense, 5'-ATGGTTTCTCAGTCCATGG-3'; nt 1766–1787) and NG697 (antisense, 5'-CAGATCAATAGAGTCTGGC-3'; nt 562–583) were used for the first-round PCR. Primers NG698 (sense, 5'-CATATGATACCTGGGAACTAGC-3'; nt 1852–1873) and NG699 (antisense, 5'-TCTTACCTTCCTTCAGGCCGTCGTT-3'; nt 506–527) were used for the second-round PCR: nucleotide numbers are in accordance with the SAV1 isolate. To amplify the 1129 nt SAV2 sequence, primers NG702 (sense, 5'-GGAGGTTACAGGCGCCTTGC-3'; nt 2205–2224) and NG701 (antisense, 5'-AACTGT-TGGCAGGCAACCTC-3'; nt 730–751) were used for the first-round PCR. Primers NG716 (sense, 5'-ACAGCCCTTCAAAGAAATCAACC-3'; nt 2228–2249) and NG703 (antisense, 5'-GGTGGATCTGGAAGGTGGTGCGC-3'; nt 702–721) were used for the second-round PCR: nucleotide numbers are in accordance with the SAV2 isolate. Nested PCR was carried out using Takara La Taq with GC buffer (TaKaRa Bio) as described previously (Okamoto et al., 1999a). The nucleotide sequence of the amplification products was determined on both strands by using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) or DYEnamic ET Terminator Cycle Sequencing kit (GE Healthcare) directly or after cloning into pT7BlueT-Vector (Novagen) or M13 phage vector (New England BioLabs). Sequence analysis was performed using GENETYXver.8 (Software Development) and OEND version 1.1.1 from the DNA DataBank of Japan (DDBJ; National Institute of Genetics, Mishima, Japan) (Ina, 1994). Sequence alignments were generated by the DDBJ version of CLUSTAL W (Thompson et al., 1994). Phylogenetic trees were constructed by using the neighbour-joining method (Saitou & Nei, 1987). The reliability of the phylogenetic results was assessed using 1000 bootstrap replicates (Felsenstein, 1985). The final tree was obtained using the TREEVIEW program (version 1.6.6) (Page, 1996).

Surprisingly, 1.9 kb PCR amplicons were exclusively obtained from four samples (MD1-032, MD1-073, MD1-160 and MD1-165) for SAV1; the amplicons were 1.0 kb longer than expected and were 92.5–99.5 % identical to SAV1 within the overlapping regions (regions x and β) (Fig. 1a). The MD1-032, MD1-073, MD1-160 and MD1-165 isolates shared identities of 90.8–99.5 % within region x (568 nt) and 94.9–99.6 % identities within region β (313 nt). Similarly, 1.7 kb PCR amplicons were obtained from two other samples (MD2-013 and MD2-099) for SAV2; the amplicons were 0.6 kb longer than expected and were 77.8–78.7 % identical to SAV2 within the overlapping regions (regions γ and δ) (Fig. 1b). The MD2-013 and MD2-099 isolates were 97.0 % identical to each other within region γ (751 nt) and 96.2 % similar to each other within region δ (336 nt). Of note, we could not obtain any other amplicons of the same size range as the ones reported by Jones et al. (2005) from the 218 subjects studied by either SAV1- or SAV2-specific PCR.

To characterize the MD1-032, MD1-073 and MD2-013 isolates over the entire genome, the genomic region that overlapped the previously amplified region at both ends was amplified by using PCR with inverted primers, with a sequence unique to each isolate (Supplementary Table S1 available in JGV Online), and the amplicons were subjected to sequence analysis. The MD1-032, MD1-073 and MD2-013 isolates had a circular genomic structure with a genomic DNA length of 3245, 3242 and 3253 nt, respectively, which were shorter than TTV and longer than TTMV. However, similar to TTV and TTMV, each isolate possessed four major ORFs (Fig. 2a), regions with a high G+C content, a coding region defined as the sequence between the beginning of ORF2 and the end of ORF4 with a high degree of genetic divergence, and a non-coding region between the end of ORF4 and the beginning of ORF2 within a relatively conserved area (Fig. 2b). ORF1 in the MD1-032, MD1-073 and MD2-013 isolates encoded a sequence of 673–677 aa that was rich in Arg at its N terminus and ORF2 encoded the conserved motif (W-X7-H-X3-C-X1-C-X5-H), both of which are highly characteristic of TTV and TTMV (Hijikata et al., 1999; Okamoto et al., 2000b;
Takahashi et al. (2000b) as well as SAV (Andreoli et al., 2006). Although the genomic organization and characteristic sequences were highly conserved, the MD1-032, MD1-073 and MD2-013 isolates differed from each other by 9.8–28.0% over the entire genome. Upon comparison of ORF1 between MD1-073 and MD2-013, low identities of 61.4% in the nucleotide sequence and 44.8% in the amino acid sequence were noted. In ORF2, these two isolates shared 74.0% similarity in the nucleotide sequence and 63.1% similarity in the amino acid sequence. The MD1-073 and MD1-032 isolates had 98.7 and 88.8% similarities to SAV1 within the overlapping 2249 nt sequence, and were 993 and 996 nt longer than SAV1, respectively. The MD2-013 isolate showed 76.2% identity to SAV2 within the overlapping 2635 nt sequence, and was 618 nt longer than SAV2. These results suggest that SAV1 and SAV2 are both deletion mutants of a novel TTV- and TTMV-like DNA virus with a highly divergent genome of 3.2 kb, which was provisionally designated ‘torque teno midi virus’ (TTMDV) in the present study as an intermediate between TTV and TTMV of 3.8–3.9 and 2.8–2.9 kb in terms of genomic length.

To elucidate the genetic relatedness of TTMDV with TTV and TTMV, a phylogenetic tree was constructed by using the neighbour-joining method (Saitou & Nei, 1987) based on the entire nucleotide sequence of ORF1 (Fig. 3a). The tree revealed that the MD1-032, MD1-073 and MD2-013
Fig. 3. Phylogenetic trees constructed based on the entire nucleotide sequence of ORF1 (a), and on the amino acid sequences of ORF1 (b) and ORF2 (c) by using the neighbour-joining method (Saitou & Nei, 1987). The tree in (a) includes the three TTMDV isolates obtained in the present study as well as 81 TTV and 13 TTMV isolates from humans and chimpanzees, whose nucleotide sequence data were retrievable from the GenBank/EMBL/DDBJ databases as of January 2007. The trees in (b) and (c) include the three TTMDV isolates obtained in the present study as well as six representative TTV and three representative TTMV isolates from humans and chimpanzees and anelloviruses from macaques (Mi-TTV3 and Mi-TTV9), douroucouli (At-TTV3), tamarin (So-TTV2), tupaias (Tbc-TTV14), pig (Sd-TTV31), dog (Cf-TTV10) and cat (Fc-TTV4). The percentages of bootstrap values generated from 1000 samplings of the data are shown near the nodes. Bar, represents the number of nucleotide or amino acid substitutions per position.
isolates segregated into the same cluster and were clearly separate from all reported TTVs and TTMVs of human and chimpanzee origin whose entire ORF1 sequence is known (Fig. 3a), suggesting that TTMDV is an independent species in the genus Anellovirus. The trees constructed based on the entire amino acid sequence of ORF1 or ORF2 also revealed that TTMDV is a virus species that is phylogenetically distinguishable not only from TTVs and TTMVs of humans and chimpanzees but also from TTVs of macaque, douroucouli, tamarin, tupaia, dog, cat and pig whose entire genomic sequence has thus far been reported (Okamoto et al., 2000a, b, 2001b, 2002) (Fig. 3b, c).

In an attempt to further examine the presence of TTMDV in the general population of Japan and to confirm the genomic length of TTMDV among isolates other than the MD1-032, MD1-073 and MD2-013 isolates obtained in the present study, inverted-nested primers that were derived from a highly conserved area among TTMDVs and SAVs but not among TTVs and TTMVs were designed derived from a highly conserved area among TTMDVs of macaque, douroucouli, tamarin, tupaia, dog, cat and pig whose entire genomic sequence has thus far been reported (Okamoto et al., 2000a, b, 2001b, 2002) (Fig. 3b, c).

In an attempt to further examine the presence of TTMDV in the general population of Japan and to confirm the genomic length of TTMDV among isolates other than the MD1-032, MD1-073 and MD2-013 isolates obtained in the present study, inverted-nested primers that were derived from a highly conserved area among TTMDVs and SAVs but not among TTVs and TTMVs were designed derived from a highly conserved area among TTMDVs and SAVs but not among TTVs and TTMVs were designed for specific amplification by PCR of the full-length circular TTMDV genome [Supplementary Fig. S1(a) available in JGV Online]. Primers used for the first-round PCR were NG574 (sense, 5′-GGACGGAGGGTCCCGAGTGG-3′; nt 86–108) and NG575 (antisense, 5′-CTGGATCCGTCCTCGACCGTC-3′; nt 63–85) and those for the second-round PCR were NG578 (sense, 5′-TGGGCGGAGGCCCCAGTTGAGT-3′; nt 113–135) and NG579 (antisense, 5′-CCGTTCATCCAGGGGTAAAATCGAC-3′; nt 45–67); nucleotide numbers are in accordance with the MD1-073 isolate of 3242 nt obtained in the present study. Amplicons were obtained from serum samples from 28 (35.9%) of 78 apparently healthy individuals in Japan, all of whom were known to be co-infected with TTV and TTMV. After electrophoresis on an agarose gel to estimate the genomic size, the amplification products of 28 PCR-positive samples migrated to the position of 3.2 kb, which is clearly longer than SAV DNAs of 2249 or 2635 nt and is comparable to TTMDV DNAs of 3242–3253 nt [Supplementary Fig. S1(b) available in JGV Online]. These results demonstrate the presence of TTMDV of 3.2 kb in the general population of Japan, who are frequently co-infected with TTV and TTMV. Since the selection of PCR primers and the length of the genomic region for PCR amplification crucially influence the detection of TTV, which has a markedly divergent genome (Itoh et al., 1999; Okamoto et al., 1999b, 2001a), the actual prevalence of TTMDV DNA needs to be developed in order to clarify the exact prevalence of TTMDV in various populations.

In conclusion, the present study revealed the presence of a novel species of anellovirus with a highly divergent genomic DNA of 3.2 kb, which was tentatively designated torque teno midi virus (TTMDV) and whose genomic length was between those of TTV and TTMV. Incapability of fishing for defective/rearranged genomes of TTMDV in the present study suggests that the SAV1 and SAV2 genomes might have been identified as an artefact. Further studies are needed to clarify the extent of genomic variability for a more precise definition of the taxonomic position of TTMDV within the genus Anellovirus, the disease associations or disease-inducing potential, as well as the virological significance of co-infection of three human anelloviruses with circular genomes of distinct lengths (2.8–2.9, 3.2 and 3.8–3.9 kb) in humans.

References


http://vir.sgmjournals.org

1943

Downloaded from www.microbiologyresearch.org by
IP: 54.70.40.11
On: Sat, 03 Aug 2019 22:52:01


