Determination and phylogenetic analysis of partial sequences from TT virus isolates

Philippe Biagini,1 Pierre Gallian,1 Houssam Attoui,1 Jean-François Cantaloube,1 Philippe de Micco1, 2 and Xavier de Lamballerie2

1 Laboratoire de biologie moléculaire, Etablissement de Transfusion Sanguine Alpes-Provence, 149 Bd Baille, 13392 Marseille cedex 05, France
2 Laboratoire de virologie moléculaire, tropicale et transfusionnelle, Faculté de Médecine de Marseille, 27 Bd Jean Moulin, 13385 Marseille cedex 05, France

Sera from French in-patients were tested for the presence of the TT virus (TTV) genome using PCR and degenerate primers located in ORF1. Thirty-six sequences were determined and compared with those deposited in databases, revealing a high degree of genetic variability between TTV isolates (up to 47% for amino acid sequences). Phylogenetic analysis demonstrated the existence of three main groups corresponding to the previously described genotypes 1 and 2 and to a new genotype 3. Isolates could be assigned to distinct genotypes if their genetic distance was > 27%. No comparable genetic criteria were found for the definition of sub-types in the region studied. A 15–31 month follow-up of three haemodialysis patients proved the existence of chronic infection by TTV. In one patient, two strains belonging to different genotypes were detected at the same time. Sequences of both ORF1 and ORF2 remained unchanged for a given strain during the follow-up.

Recently, TT virus (TTV) has been identified in sera of patients with post-transfusion hepatitis showing elevated transaminases (Nishizawa et al., 1997). Preliminary studies have shown the viral genome to be a single-stranded DNA approximately 3.7 kb long, harbouring two putative open reading frames (ORFs). ORF1 (nt 589–2898) and ORF2 (nt 107–702), would encode, respectively, proteins consisting of 770 and 202 amino acids (Okamoto et al., 1998). This structural organization, the multiple TATA sequences and the polyadenylation signals suggest that TTV is a member of the family Parvoviridae. Accordingly, ORF1 would encode a structural protein and ORF2 a non-structural one. However, none of the known TTV sequences exhibit any significant homology to those of paroviruses.

A previous study described two viral genotypes and four sub-types for Japanese isolates belonging predominantly to type 1 (Okamoto et al., 1998). The virus is widely distributed in populations with parenteral risk exposure: haemodialysis patients (46%), intravenous drug users (19–40%) and haemophiliacs (27–48%) (Biagini et al., 1998; Okamoto et al., 1998; Simmonds et al., 1998). TTV was also detected at a lower prevalence in voluntary blood donors (1–9%). Although it was observed that the prevalence of TTV infection was high in patients with liver disease, i.e. hepatocellular carcinoma (39%), non-A to G liver diseases (46%), non-A to G fulminant hepatitis (47%) and cirrhosis (48%), TTV could not be formally identified as an aetiological agent for a specific human pathology.

We report here 36 partial TTV sequences from European isolates. An analysis of the variability and phylogenetic distribution for these TTV sequences and those retrieved from databases was performed. We also report the results of a study of the natural history of viral infection over a period of 31 months.

Blood samples from patients hospitalized in various healthcare units located in Marseilles (France) were collected under nuclease-free conditions and sera were stored at −80 °C. Our population included three haemodialysis patients whose sera were recovered over a period of 1 to 3 years.

Viral DNA extracted from 75 µl of serum using the High Pure Viral Nucleic Acid kit (Boehringer Mannheim) was submitted to two rounds of PCR amplification under standard conditions (Biagini et al., 1997) at an annealing temperature of 50 °C. Degenerate oligonucleotide primers derived from those described by Okamoto et al. (ORF1) were designed for the amplification of the most divergent variants described to date (Table 1). An additional primer, RD037d, was used to amplify a sequence specific to type 1. Primers used for amplification in ORF2 were designed from the previously published sequence

Author for correspondence: Xavier de Lamballerie.
Fax + 33 491 18 95 98. e-mail virophdm@lac.gulliver.fr
The GenBank accession numbers of the sequences reported in this paper are AF083134–AF083168.
Fig. 1. Amino acid substitutions in protein sequences of TTV strains (ORF1 region; amino acids 453–522). Underlined sequences represent highly conserved amino acid positions; type-specific positions are boxed.
of a type 1 isolate (accession no. AB008394). Amplicons were revealed by ethidium bromide staining following agarose gel electrophoresis. PCR products were purified with a Qiagen PCR Purification kit and directly sequenced using second-round primers, the DNA D-Rhodamine Sequencing kit and an ABI-Prism 377 sequence analyser (both from Perkin Elmer). Where co-infection was suspected, amplicons were cloned into the pGEM-T vector (Promega), transfected into *E. coli* XL-Blue competent cells and sequenced on both strands using universal M13 primers.

Sequence alignments were generated by the CLUSTAL W version 1.74 program (Thompson *et al*., 1994). Phylogenetic analysis of aligned sequences was performed using the software program MEGA (Kumar *et al*., 1993).

TT virus was first detected in the sera of patients with non-A to G hepatitis following blood transfusion. The first data from Nishizawa *et al.* (1997) suggested that patients could recover from infection after a few weeks. By contrast, Simmonds *et al.* (1998) suggested that the virus was able to establish a chronic infection, since haemophiliacs were found to be infected 10 years after their last exposure to non-virally inactivated clotting factor concentrates.

In this study, serum samples from three haemodialysis patients, HDS1, HDS2 and HDS3, without markers for hepatitis A to G viruses, were retrospectively tested for the presence of TTV using PCR primers located in ORF1. The virus was detected in sera of all patients, during a follow-up of 15 months for patient HDS1, 23 months for patient HDS2 and 31 months for patient HDS3. This demonstrated the existence of chronic infection by TTV. Molecular analysis of viral strains showed that patient HDS2 was infected by a type 1b strain, and patient HDS3 by a type 2 strain, suggesting that chronic infection may not be restricted to a particular genotype.

The hypothesis of a re-exposure to viral infection leading to an apparent chronic infection could be excluded by the high degree of sequence variability that was found between isolates (up to 47%), in contrast to the absolute conservation of viral sequences in a given infected individual. The absence of changes of viral sequences could be verified for all three patients in ORF1 (nt 1945–2154), and in ORF2 (nt 210–462) for patient HDS1. This patient (infected by a type 1a strain) was the only one who gave positive PCR results using primers located in ORF2. Nine positions in the amplified sequences were found to be altered (3.6%) leading to seven amino acid changes (8.3%), by comparison with the reference sequence AB008394 (type 1a). The analysis of eight recently reported Chinese sequences (AB011486–AB011491, AB011493, AB011494) belonging to types 1a and 1b revealed mutations specific for type 1a at the position of our reverse primers, which explains the negative results obtained for patients HDS2 and HDS3 (types 1b and 2, respectively).

Patient HDS1 was shown to have been infected initially by a type 1a strain. During the first months of 1998, a strain belonging to type 2 was also detected in this patient’s serum.

### Table 1. Oligonucleotide primers used in amplification of TTV strains

<table>
<thead>
<tr>
<th>Region</th>
<th>First-round primers</th>
<th>Second-round primers</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>NG059d: WCAGACAGAGGMGAAGGMAAYATG (1900–1923)</td>
<td>NG063d: CTGGCATYTTWCCRTTTCCAAART (2162–2185)</td>
<td>271 nt</td>
</tr>
<tr>
<td></td>
<td>NG061d: GGMAAYATGYTRTGGATAGACTGG (1915–1938)</td>
<td>NG063d: CTGGCATYTTWCCRTTTCCAAART (2162–2185)</td>
<td>178 nt</td>
</tr>
<tr>
<td></td>
<td>RD037d: GCAGCAGCATATGGWTATKT (2008–2027)</td>
<td>NG063d: CTGGCATYTTWCCRTTTCCAAART (2162–2185)</td>
<td>295 nt</td>
</tr>
<tr>
<td>ORF2</td>
<td>TTVNS1: GGGTGCCGAAGGTGAGTTTAC (175–195)</td>
<td>TTVNS2: GCGGGGCACGAAGCACAGAAG (474–494)</td>
<td>295 nt</td>
</tr>
<tr>
<td></td>
<td>TTVNS3: AGTTTACACACCGAAGTCAAG (195–215)</td>
<td>TTVNS4: AAC4CACAGAAGCAAGATG (463–483)</td>
<td>271 nt</td>
</tr>
<tr>
<td></td>
<td>TTVNS5: AGCACAGAAGCAAGATGATTA (463–483)</td>
<td>TTVNS6: GCCGCGGCGCAGAAGAGAAG (474–494)</td>
<td>295 nt</td>
</tr>
</tbody>
</table>
Fig. 2 (a). For legend see facing page.
At the end of a 15 month follow-up, both strains could still be detected without any alteration in their respective nucleotide sequences. Two hypothesis can therefore be formulated: (i) patient HDS1 was co-infected by two different viral strains since the beginning of our study, but the type 2 strain could not be detected in the initial samples, probably because of a very low viral load; (ii) this patient was first infected by a type 1a strain, and became superinfected by a type 2 strain.

Thirty-six sequences (designated TTm1–30 and HDSx, and characterized in this study) were compared to 51 TTV sequences retrieved from GenBank (nt 1945–2154, ORF1). Out of the 210 nucleotide positions analysed, 135 (64.3%) were found to be modified at least once with respect to the reference sequence AB008394. No insertions or deletions were observed. Among the 70 amino acids encoded, 48 positions (68.6%) were found to be modified at least once (Fig. 1). However, three amino acid motifs were conserved (positions 463–466, SKCL; 471–474, PLWA; 500–503, IRSP). Comparisons of these motifs with a database containing 5380 identified functional motifs showed no significant matching (Blocks searcher version 10.1; on-line at www.blocks.fhcrc.org).

Comparison of TTV sequences with sequence databases did not identify any significant identity at either nucleotide or amino acid levels.

Pairwise comparison of amino acid sequences was performed to assess the genetic relationships between the 87 TTV isolates tested, using the p-distance algorithm for distance determination, the neighbour-joining method for tree-drawing and a bootstrap resampling of 500 replications. An unrooted phylogenetic tree (Fig. 2a) was constructed, and showed a similar distribution to that obtained from the corresponding nucleotide sequences. Three main groups supported by a 100% bootstrap confidence level (BCL) could be distinguished. (i) Group 1 includes all strains related to type 1 isolates (according to Okamoto et al., 1998). Sub-groups (corresponding to sub-types 1a and 1b) can be distinguished with 54% and 31% BCL respectively. Only seven isolates, collected in Marseilles, belong to this group. (ii) Group 2 includes all strains related to type 2 isolates. Isolates previously assigned to sub-types 2a and 2b are located in two different clusters. Three isolates from Marseilles cluster within sub-type 2b and none within sub-type 2a. Nine isolates from Marseilles constitute a third sub-group together with three strains from Germany and one from Japan. Twelve strains from Marseilles form a new sub-group supported by an 86% BCL value. The German isolate AF060550 is located at an intermediate position between the isolates described above and those belonging to group 3. (iii) Three isolates from Marseilles constitute a third and new phylogenetic group.

In agreement with these results, the amino acid alignment showed the existence of patterns specific to each group (Fig. 1). A histidine residue was present at position 493 for the 45 isolates belonging to group 1. A tyrosine residue was present at position 519 in the 39 isolates of group 2 (except for isolate AF060550). A methionine residue was found in position 509 in the three isolates of group 3.

The important variability in amino acid sequences (up to 47%) between the different TTV isolates, and the perfect conservation of sequences in both ORF1 and ORF2 in the three infected patients during their follow-up, suggest that the current TTV variants are separated from their common phylogenetic ancestor by a considerable distance in time. This implies a long history of infection in humans.
Data from the pairwise comparison of amino acid sequences were used to construct a p-distance/frequency histogram (Fig. 2b). Among the different cut-off values tested for the definition of genotypes, only one (27%) gave results that matched our phylogenetic findings. Accordingly, two isolates could be assigned to distinct genotypes when the genetic distance between them (amino acid sequences, positions 453–522) was > 27%. According to this definition, isolates from phylogenetic group 1 would constitute genotype 1, isolates from phylogenetic group 3 would constitute genotype 3 and isolates from phylogenetic group 2, including the AF060550 strain, would constitute genotype 2. If isolate AF060550 is considered to be a distinct genotype, the genetic distance between that genotype and some isolates of genotype 2 would be lower than that between numerous isolates from genotype 2. This would ruin any attempt to define genotypes on the basis of genetic distances in this genomic region.

The definition of genotypes is illustrated in Fig. 2(c), which displays a simplified ultrametric tree constructed using the UPGMA method. Analysis of the pairwise distribution matrix showed that the genetic distance was 31–47% between types 1 and 2, 35–44% between types 1 and 3 and 27–39% between types 2 and 3; the variability within genotypes was 0–17% for type 1, 0–26% for type 2 and 1–6% for type 3.

It was not possible to identify a cut-off value permitting the grouping of isolates in distinct sub-types on the basis of amino acid comparison. When using nucleotide sequences, such a cut-off value could be defined within genotype 1 only (two isolates would belong to different sub-types if the genetic distance was > 8%). When applied to type 2, this definition produces subgroups that do not reflect the effective phylogenetic distribution. It is therefore not possible to establish a consensual definition of sub-types in that particular genomic region based on genetic distances.

The ORF1 genomic region of TTV may, by analogy with the Paroviridae genomic structure, correspond to the capsid-coding sequence of the virus. Previous studies on the variability of the structural region among various parvovirus B19 isolates showed 0–4.2% amino acid variability (Erdman et al., 1996; Hemauer et al., 1996). Comparison between the Japanese reference sequence (AB008394) and eight Chinese sequences (AB011486–AB011491, AB011493, AB011494) showed up to 12.9% amino acid variability for the full-length ORF1 sequence; similar results are found when studying the partial sequence spanning amino acids 453–522 (this study). The degree of variability of ORF1 is therefore much more important in the 87 TTV isolates studied (> 45%) than in parvovirus B19 isolates (< 5%).

Additional studies are required to gain insight about the structural and clinical aspects of TTV. The determination of the 5’ and 3’ ends (which are still unanalysed) will permit the establishment of its genomic organization, and therefore its taxonomic status. Its involvement in human pathology and the relationships with the host immune system remain unclear and further investigations will be necessary to determine whether TTV is an orphan virus or a real pathogen. As a first step in the description of the natural history of viral infection, this study demonstrates the existence of chronic forms of infection and the possibility of co-infection by viruses belonging to different genotypes.

The authors which to thank R. Khayat de Chesse for excellent technical assistance.

References


Received 7 August 1998; Accepted 28 September 1998