A new group of transmissible single-stranded (ss) DNA viruses (SENV) distantly related to the large TT virus (TTV) family was recently identified. Eight different SENV isolates have been found, some with an association with posttransfusion hepatitis. A phylogenetic analysis of near-complete open-reading frame 1, including conserved motifs and excluding recombinant regions, was performed. The analysis used TTV-like minivirus as an outgroup, to determine a root of the phylogenetic tree, and compared 8 SENV isolates, 6 prototype TTV isolates, and 7 TTV variants (including SANBAN, TUS01, PMV, and YONBAN). Four distinct clusters separated by a bootstrap value of 100% were observed. YONBAN isolates formed a distinct outer group, representing the earliest recognized phylogenetic divergence (group 1). Prototype TTV formed group 2, PMV formed group 3, and SENV, SANBAN, and TUS01 isolates formed group 4, the most recently evolved group. This taxonomic classification suggests that these circular ssDNA viruses probably evolved from a common ancestor virus.

Indeed, the sequence data on distant related members of the TTV family [6–14] have grown greatly in recent years, and the classification systems [6, 7, 9–11] have increased in complexity. In addition, Worobey [15] recently showed that TTV genomes possess recombination breakpoints that are more common in the noncoding region than in the coding region. Thus, a comprehensive and logical overview of the superfamily of these ssDNA viruses is needed. In this study, we selected sequences of the entire open-reading frames (ORF1–ORF3) known to have few recombinant regions [15] for molecular evolutionary analysis. TTV-like minivirus (TLMV) [16] was used as an outgroup, to determine a root of the phylogenetic tree and to provide a possible evolutionary framework to the taxonomic classification. Sequences of the entire ORF1–ORF3 were compared for 8 novel SENV isolates [3, 4], the prototype TTV isolates [2, 12–14], and several TTV variants, including SANBAN [8], TUS01 [9], PMV [11], and YONBAN [17]. This approach provided a clear classification of the diverse agents in this complex family and a perspective on their interrelationship.

Materials and Methods

**Serum samples.** Human serum samples known to contain SENV-D DNA or SENV-H DNA were used as the source of virus for cloning and sequencing of the genome. These samples were stored at −70°C before use.

**Detection of SENV DNA.** Nucleic acids were extracted from 100 μL of serum by using QIAamp DNA Blood Mini Kit (Qiagen). The DNA was recovered in 160 μL of elution buffer containing 10 mM Tris-HCl and 0.5 mM EDTA (pH 9.0). SENV-D DNA and SENV-H DNA were determined by the polymerase chain reaction (PCR) with SENV-specific primers, as described by Primi et al. [4].
Inverted PCR

**Figure 1.** Putative physical maps of SEN-D and SEN-H viruses. Inverted polymerase chain reaction (PCR) with seminested primers derived from conserved region between TT virus and SEN virus (SENV) was done. First-round PCR was done with sense primer SENs1 and antisense primer NG132; second-round PCR was done with sense primer SENs1 and antisense primer SENas1. SENV-D and SENV-H are circular and are composed of 3788 and 3815 nt, respectively, including >3 putative open-reading frames (ORFs) overlapping ORF2 and ORF1 and ORF3 at the 3' end of ORF1. aa, Amino acid; nt, nucleotides.

For SENV-D, sense primer D10S (5'-GTAACTTTGCGGTCACTGCC-3') and antisense primer LUCKY 2AS (5'-CCCTGGTTKSAAAKGTYTGATAGT-3'; K = G or T, S = C or G, Y = C or T) were used in 50 μL of PCR mixture containing 1.75 U of Perkin-Elmer AmpliTaq DNA polymerase (Roche Molecular Systems). The PCR procedure involved 40 cycles consisting of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min for each cycle, followed by the extension reaction at 72°C for 9 min in a thermal cycler (model 9700; Perkin Elmer). For SENV-H, sense primer C5S (5'-GGTGCCCCTWGTYAGTTGGCGGTT-3'; W = A or T, Y = C or T) and antisense primer LUCKY 2AS were used, and PCR cycles as described above were used, except that the annealing temperature was at 62°C for 1 min. PCR products were analyzed by DNA EIA (Diasorin). This involved the hybridization of amplified DNA with SENV-D-specific 5' biotinylated probe (5'-ATGATAGGCTTCCYTTTAACTATAACCCA-3'; Y = C or T) or SENV-H-specific 5' biotinylated probe (5'-CCCCCTCCAGGTATTGCATGAAGTATTAC-3'). Specimens with an optical density >0.350 were considered to be positive for SENV-D or SENV-H.

**Genomic extension.** To examine the possibility that the SENV genome was circular, inverted PCR with seminested primers derived from the conserved region between TTV and SENV was performed with TaKaRa LA Taq and GC buffer II (Panvera) with SENV-D-positive or SENV-H-positive samples. As shown in figure 1, the first-round PCR was performed for 40 cycles (94°C for 1 min, 55°C for 1 min, and 72°C for 2 min for each cycle) with a sense primer (SENs1: 5'-AGACTCCAGTTGCCATGG-3') and an antisense primer (NG132: 5'-AGCCCGAATTGCCCCTTGAC-3'). The second-round PCR was performed for 40 cycles (94°C for 1 min, 60°C for 1 min, and 72°C for 2 min for each cycle) with a sense primer (SENs1) and an antisense primer (SENas1: 5'-ACTCAGCCATTC-GGAAGTG-3'). All PCR products were cloned into pCR2.1-TOPO vector (Invitrogen), and 5 clones were sequenced. Sequencing reactions were done with Prism dGTP Big Dye (Perkin-Elmer Applied Biosystems) in the ABI 377 DNA automated sequencer, according to the manufacturer's protocol.

**Computer analysis of sequences.** The DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory/GenBank accession numbers of TTV, SANBAN, YONBAN, TLMV, and SENV isolates used in this study are as follows: complete sequences, AB017610 (TA278) [2], AF122914 (JA20) [12], AF122918 (JA2B) [12], AF247138 (T3PB) [13], AB025946 (SANBAN) [8], AB038619 (TTVsan-IR1031), AB017613 (TUS01) [9], AF247137 (TUPB) [13], AF261761 (PMV) [11], AB038622 (TTVyon-LC011), AB038623 (TTVyon-KC186), and AB038631 (TLMV-NLC030); and near-complete sequences, AB030486 (JaCHC2c) [14] and AB030488 (JaBD89) [14]. Eight SENV sequences (A–H) were obtained from the patent [4]: AX025667 (SENV-A), AX025677 (SENV-B), AX025718 (SENV-C), AX025730 (SENV-D), AX025761 (SENV-E), AX025822 (SENV-F), AX025830 (SENV-G), and AX025838 (SENV-H). Accession numbers for the remainder of sequences (including guanine and cytosine-rich regions) of SENV-D and SENV-H are AB048602 and AB048603, respectively. The sequences were aligned with Clustal W (version 1.8; DDBJ) and were adjusted in correspondence to the motifs by using the GCG program (Wisconsin Package, version 10.0).

We used ODEN software (version 1.1.1; DDBJ) [18] with the 6-parameter method [19], to determine the number of nucleotide substitutions per site (genetic distance) between isolates. On the basis of these values, a phylogenetic tree was constructed by the neighbor-joining method [20]. We used a program from DDBJ with the midpoint rooting option to plot the tree. To confirm the reliability of the phylogenetic tree, bootstrap resampling tests were done 1000 times [21]. Frequency distribution analysis was per-
Figure 2. Amino acid (aa) alignment of putative open-reading frame (ORF) 1. Conserved sequence motifs in putative replication-associated proteins (Rep proteins) mediate rolling-circle DNA replication of SEN virus (SENV) family. Two conserved motifs (1 and 3 for Rep proteins) and conserved ATP/GTP-binding motif 4 (p-loop) are in boldface above aligned sequences. Putative highly conserved protein kinase phosphorylation sites are in boldface. TYR, tyrosine kinase phosphorylation site \[[RK]-x(2,3)-[DE]-x(2,3)-Y]\]; CK2, casein kinase 2 phosphorylation site \[[ST]-x(2)-[DE]\]; PKC, protein kinase C phosphorylation site \[[ST]-x-[RK]\]. An "x" indicates no consensus in this position. Nos. indicate aa position in SENV-C sequences.

Results

Genome maps. There are ≥8 different isolates of SENV (A–H). Inverted PCR with upstream antisense primers and downstream sense primers generated ~600-bp products representing the rest of the genome (figure 1). The SENV-D genome of 3788 nt and SENV-H genome of 3815 nt are slightly shorter than the prototype TTV genome (TA278; 3853 nt). SENV had ≥3 ORFs: overlapping ORF1 and ORF2 and a novel ORF3 at the 3’ end of ORF1 (figure 1).

Deduced aa sequences encoded by ORF1. The ORF1 from the 8 SENV isolates in this study and related ssDNA virus isolates are shown in figure 2. The aa residues encoded by ORF1 of SENV-A, SENV-B, SENV-C, SENV-D, SENV-E, SENV-F, SENV-G, and SENV-H consisted of 642, 679, 753, 754, 743, 758, 763, and 762 aa, respectively. In the N-terminal region of ORF1 was an arginine/lysine (Arg/Lys)–rich domain, highly conserved among most SENV, as well as among prototype TTV, TTV variants, and chicken anemia virus (CAV), although a corelike immunogenic region of SENV-A and SENV-B was absent (data not shown). TTV possessed 2 of 4 conserved motifs (motifs 1 and 3) present in the putative replication-associated proteins (Rep proteins), which are involved in rolling-circle replication [22] similar to many plant and animal circoviruses [23, 24]. The largest ORF1 of SENV also possessed the 3 motifs of Rep proteins. As shown in figure 2, motif 3 was highly conserved among all SENV isolates, and motifs 1, 2, and 4 (p-loop) were also conserved among some isolates. Thus, these conserved regions were used for phylogenetic analyses.

Analysis of the ORF1 of SENV-A, SENV-B, SENV-C, SENV-D, SENV-E, SENV-F, SENV-G, and SENV-H contained 2, 1, 2, 0, 4, 3, and 2 potential N-glycosylation sites; 3, 3, 6, 3, 3, 1, 4, and 5 tyrosine kinase phosphorylation sites; 9, 5, 9, 6, 11, 7, 12, and 11 casein kinase 2 phosphorylation sites; 2, 1, 4, 3, 1, 3, and 2 amidation sites; and 2, 5, 4, 5, 7, 7, 4, and 6
N-myrstoylation sites, respectively. The ORF1 of SENV-A, SENV-C, and SENV-H at the C-terminus of the sequences (positions 617–638, 729–750, and 738–759, respectively) contained the leucine zipper pattern (LQLVMFQLSR) which was absent in the ORF1 of TTV.

As shown in table 1, the aa homology of ORF1 among 8 SENV isolates was 39.1%–76.8%. The alignment of the aa sequences of ORF1 of all SENV isolates (A–H), compared with that of prototype TTV (TA278), showed sequence sharing between 33.1%–36.5% (table 1). Likewise, when we compared SENV isolates with SANBAN, TUS01, PMV, and YONBAN, sequences sharing 39.3%–60.5%, 46.1%–78.6%, 34.2%–37.1%, and 29.1%–36.1%, respectively, were obtained. Of interest, the aa sequences sharing 39.3%±60.5%, 46.1%±78.6%, 34.2%±37.1%, and 29.1%±36.1%, respectively, were obtained. Of interest, the sequences sharing 39.3%±60.5%, 46.1%±78.6%, 34.2%±37.1%, and 29.1%±36.1%, respectively, were obtained.

### Table 1. Pairwise percent homology of nucleotide and amino acid sequences within entire open-reading frame 1 in 8 SEN virus (SENV) isolates and other related virus isolates (SANBAN, TUS01, TTV, PMV, YONBAN, and TLMV).

<table>
<thead>
<tr>
<th>Virus isolate</th>
<th>SENV-A</th>
<th>SENV-B</th>
<th>SENV-C</th>
<th>SENV-D</th>
<th>SENV-E</th>
<th>SENV-F</th>
<th>SENV-G</th>
<th>SENV-H</th>
<th>SANBAN</th>
<th>TUS01</th>
<th>TTV</th>
<th>PMV</th>
<th>YONBAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SENV-B</td>
<td>56.0/55.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENV-C</td>
<td>63.4/52.6</td>
<td>60.7/48.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENV-D</td>
<td>66.4/57.3</td>
<td>61.2/51.0</td>
<td>63.5/52.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENV-E</td>
<td>58.9/47.8</td>
<td>59.1/43.7</td>
<td>58.2/46.6</td>
<td>60.0/49.2</td>
<td>52.3/39.1</td>
<td>54.6/49.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENV-F</td>
<td>67.2/57.0</td>
<td>60.9/50.7</td>
<td>61.1/49.2</td>
<td>77.6/67.8</td>
<td>59.4/47.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENV-G</td>
<td>56.6/53.5</td>
<td>60.0/47.9</td>
<td>54.7/46.8</td>
<td>54.8/49.2</td>
<td>52.3/39.1</td>
<td>54.6/49.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENV-H</td>
<td>57.5/51.3</td>
<td>59.9/48.1</td>
<td>57.8/46.9</td>
<td>57.9/48.6</td>
<td>51.5/39.3</td>
<td>51.3/46.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANBAN</td>
<td>58.6/45.7</td>
<td>53.1/43.6</td>
<td>59.0/47.2</td>
<td>57.7/44.9</td>
<td>58.6/40.5</td>
<td>59.7/46.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51.3/46.1</td>
<td>51.3/46.1</td>
<td></td>
</tr>
<tr>
<td>TUS01</td>
<td>57.4/56.0</td>
<td>57.4/57.0</td>
<td>62.9/52.2</td>
<td>59.1/46.1</td>
<td>63.5/52.0</td>
<td>61.2/47.5</td>
<td>50.0/49.9</td>
<td>52.3/44.5</td>
<td></td>
<td></td>
<td>51.3/46.1</td>
<td>51.3/46.1</td>
<td></td>
</tr>
<tr>
<td>TTV</td>
<td>51.0/33.1</td>
<td>54.0/34.8</td>
<td>52.7/36.5</td>
<td>53.2/36.1</td>
<td>53.2/36.1</td>
<td>53.2/37.0</td>
<td>48.1/34.2</td>
<td>51.8/36.6</td>
<td>54.2/38.4</td>
<td>49.8/36.7</td>
<td>55.3/37.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMV</td>
<td>53.2/34.8</td>
<td>50.6/35.8</td>
<td>53.0/35.3</td>
<td>53.1/37.1</td>
<td>53.3/36.5</td>
<td>53.2/37.0</td>
<td>48.1/34.2</td>
<td>51.8/36.6</td>
<td>54.2/38.4</td>
<td>49.8/36.7</td>
<td>55.3/37.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YONBAN</td>
<td>48.1/29.1</td>
<td>49.4/30.7</td>
<td>48.8/32.0</td>
<td>49.3/32.0</td>
<td>48.0/32.6</td>
<td>49.3/32.9</td>
<td>49.3/36.6</td>
<td>48.8/30.4</td>
<td>51.5/31.9</td>
<td>49.8/28.6</td>
<td>49.2/33.3</td>
<td>48.3/32.1</td>
<td></td>
</tr>
<tr>
<td>TLMV</td>
<td>45.3/20.8</td>
<td>46.0/21.0</td>
<td>45.9/22.9</td>
<td>45.2/24.0</td>
<td>45.1/24.2</td>
<td>46.0/24.5</td>
<td>47.4/30.3</td>
<td>45.6/29.6</td>
<td>44.9/24.0</td>
<td>45.7/32.5</td>
<td>47.3/24.2</td>
<td>42.7/25.3</td>
<td>48.1/26.5</td>
</tr>
</tbody>
</table>

**NOTE:** Nos. show percent homology of nucleotide sequence/amino acid sequence. Closest homology is in boldface.

In tier B, with genetic distances of 0.72145–0.93253, com-
Figure 3. Amino acid (aa) alignment of putative open-reading frame (ORF) 3 of single-stranded DNA viruses vs. some aa sequences obtained from BLASTP search. These viruses are SANBAN (AB025946), SANBANs039 (AB038620), TUS01 (AB017613), prototype TTV (TA278 [AB017610], JA20.G1b [AF122914], JA2B.G2 [AF122918], and T3PB.G4 [AF247138]), PMV (AF261761), YONBANlc011 (AB038622), TLMVnlc030 (TTV-like minivirus, AB038631), and SENV-A to SENV-H. They have a certain homology with C-promoter binding factor±1 interacting corepressor of *Homo sapiens* (AF098297), *Caenorhabditis elegans* cosmid T28H10 (Z75551), DNA topoisomerase I protein of *Drosophila melanogaster* (AE003759), nuclear protein gar2 of *Schizosaccharomyces pombe* (S55785), nuclear protein SDK2 of *Xenopus leavis* (Y10350), nuclear protein SDK3 of *H. sapiens* (Y10351), and *Arabidopsis thaliana* chromosome 2 (AC006593). Nos. indicate aa position in SENV-C sequences.

Cluster of basic aa (R and K) is indicated in boldface; serine (S)-rich tract is boxed. Alignment shows that the C-terminus has an S-rich tract preceded by a cluster of basic aa.

Comparison of the sequences of 8 novel SENV isolates to prototype TTV, PMV, and YONBAN isolates showed such great genetic distances that they may be interpreted as different virus species beyond genotypes. In figure 4, the genetic distances between SENV-C and SENV-H, SENV-D and SENV-F, and SENV-B and TUS01 were between 0.23348 and 0.29066. Thus, these variants are intermediate between TTV subtypes and TTV genotypes. Because these sequences were closely related to each other, it is difficult to determine whether the tier represented subtype or genotype relationships between sequences. However, the aa homology of the putative ORF1 was 76.2% between SENV-C and SENV-H, 76.8% between SENV-D and SENV-F, and 78.6% between SENV-B and TUS01, compared with <70% among different genotypes of the prototype TTV. On the basis of this observation, we suggest combining SENV-C and SENV-H into one genotype, SENV-D and SENV-F into another, and SENV-B and TUS01 into yet another.

The unrooted phylogenetic tree based on these pairwise genetic distances is shown in figure 5A. On the basis of the putative entire coding region and by using all available nucleotides (2612 nt; all codon positions), 4 different clusters separated by a bootstrap value of 100% were created (figure 5A): YONBAN isolates (group 1), prototype TTV isolates (group 2), PMV (group 3), and SENV, SANBAN, and TUS01 isolates (group 4). The genetic distances among these 4 groups were mainly distributed within area B in figure 4. Group 1 consisted of YONBAN isolates recently described by Takahashi et al. [17] that were distant from group 2 (0.88240–0.93158), group 3 (0.89034–0.89048), and group 4 (0.81528–0.91752). Group 2 consisted of prototype TTV isolates including ≥6 different genotypes by use of primer sets at the TTV N22 region [7]. Eight SENV isolates formed the putative fourth cluster (group 4) and included SANBAN and TUS01 isolates. Group 4 had ≥7 genotypes separated by a bootstrap value of 100%: SENV-A, SENV-B/TUS01, SENV-C/H, SENV-D/F, SENV-G, and SANBAN isolates. SENV-B was the closest to the TUS01 isolate previously reported by Okamoto et al. [9], with a genetic distance (nucleotide homology) of 0.23061 (79.7%), which suggests that they are of the same genotype. SENV-E was closest to isolate SANBAN reported by Hijikata et al. [8], with a genetic distance (nucleotide homology) of 0.39125 (70.1%), indicating that these viruses were different genotypes.

On the basis of the great genetic distances and poor aa sequence homology among these 4 groups, we hypothesize that these groups are remotely related, but represent different viruses in a TTV-related family. To determine a root of the phylogenetic
A phylogenetic analysis of 8 different SENV isolates, prototype TTV, and some TTV variants, including SANBAN, TUS01, PMV, and YONBAN, was performed by using near-complete ORF1 sequences, with TLMV as an outgroup. This analysis revealed 4 distinct groups; group 1, novel YONBAN isolates; group 2, the prototype TTV isolates; group 3, PMV only; and group 4, SENV, SANBAN, and TUS01 isolates. The 4 groups were created within one great cluster that differed from TLMV, suggesting that prototype TTV, PMV, SENV (SANBAN and TUS01), and YONBAN probably evolved from a common ancestor. Of note, group 1 apparently had diverged from a common ancestor earlier than the other groups. YONBAN may be the origin of TTV-related species. Group 4, including SENV, appeared to be the most recently divergent group. The genetic distances, based on the frequency distribution of pairwise genetic distances produced a range of 0.72145–0.93158 among the 4 groups comprising tier B (figure 4). The genetic distances of near-complete sequences among the 4 groups were greater than those between GBV-A [27] cloned isolate and human serum samples (0.70357–0.73085). This suggests that the 4 groups are different virus species. However, such taxonomic matters should be determined by the International Committee on Taxonomy of Viruses or other special groups.

A large number of partial distinct sequences have been ob-

Figure 4. Pairwise genetic distances of entire open-reading frame (ORF) for 21 isolates, including SEN virus (SENV), prototype TT virus (TTV), and TTV variants. Genetic distances are on X-axis. Frequency of occurrence of discrete distances is on Y-axis. Tier A is composed of genetic distances of 0.33283–0.70492, which may be classified as genotypes; genetic distances among SEN isolates (0.40521–0.70492) were much greater than those among prototype TTV isolates (0.33283–0.38036). Tier B is composed of genetic distances of 0.72145–0.93158. Comparison of sequences for 8 novel SENV, prototype TTV, PMV, and YONBAN isolates used in this study show such large genetic distances distributed within tier B, which may be interpreted as different virus species. Genetic distances between SENV-C and SENV-H, SENV-D and SENV-F, and SENV-B and TUS01 are between 0.23348 and 0.29066 (*). Thus, these variants are intermediate between TTV subtypes and TTV genotypes.

Discussion

SENV belongs to the superfamily of TTV-related viruses. It is a small, most likely nonenveloped and single-stranded circular virus [3, 4]. Comparison of the genomic sequences, their encoded proteins, and biophysical characteristics of the virus may shed light on the evolutionary history of SENV in relation to TTV.

The first W of a TTV/CAV-common motif WX\textsuperscript{2}HX\textsuperscript{2}CXCX\textsuperscript{2}H was not conserved among SENV-A, SENV-C, SENV-G, and SENV-H (data not shown). pORF3 had a unique S-rich tract preceded by a cluster of basic aas (R and K). For the latter shorter proteins, pORF2 and 3, there was poor aa homology among the 4 groups. This precluded use of these proteins as candidates for meaningful comparison. Hence, to establish a suitable classification, use of aligned coding regions, including conserved motifs and excluding recombinant regions, among the 4 groups and TLMV was essential.

ORF1 aa sequences were aligned with Clustal W (version 1.8) and were adjusted in correspondence to the motifs. On the basis of the aa alignment, the coding nt alignment then was rearranged. A phylogenetic tree was constructed by using the neighbor-joining method, on the basis of the near-complete ORF1 nt alignment (1485 nt; all available codon positions) with conserved Arg/Lys-rich regions and motifs of Rep proteins, excluding putative hypervariable regions [26]. As shown in figure 5B, 4 groups (1–4) were created within one great cluster that differed from TLMV. Within this large cluster, YONBAN isolates formed a distinct outer group, indicating the first divergent species (group 1). Group 2 consisted of prototype TTV and group 3 was PMV. Finally, group 4, including the 8 SENV isolates, probably formed the last and most recent divergence from a common ancestor. Thus, prototype TTV, PMV, SENV (SANBAN and TUS01), and YONBAN probably evolved from a common ancestor.

In the analysis of group 4 including SENV in relationship to group 1 (YONBAN), group 2 (prototype TTV) and group 3 (PMV), we were confronted with a choice of 3 regions to compare: proteins encoded by ORF1 (pORF1), ORF2 (pORF2), and ORF3 (pORF3). An examination of these regions showed that only pORF1 was suitable. pORF1 possessed an N-terminal Arg/Lys-rich domain, several conserved protein kinase phosphorylation sites, and motifs of Rep proteins. pORF2 possessed a motif WX\textsuperscript{2}HX\textsuperscript{2}CXCX\textsuperscript{2}H, conserved among TTV, TLMV, and CAV. However, the first W of a TTV/CAV-common motif WX\textsuperscript{2}HX\textsuperscript{2}CXCX\textsuperscript{2}H was not conserved among SENV-A, SENV-C, SENV-G, and SENV-H (data not shown). pORF3 had a unique S-rich tract preceded by a cluster of basic aas (R and K). For the latter shorter proteins, pORF2 and 3, there was poor aa homology among the 4 groups. This precluded use of these proteins as candidates for meaningful comparison. Hence, to establish a suitable classification, use of aligned coding regions, including conserved motifs and excluding recombinant regions, among the 4 groups and TLMV was essential.

ORF1 aa sequences were aligned with Clustal W (version 1.8) and were adjusted in correspondence to the motifs. On the basis of the aa alignment, the coding nt alignment then was rearranged. A phylogenetic tree was constructed by using the neighbor-joining method, on the basis of the near-complete ORF1 nt alignment (1485 nt; all available codon positions) with conserved Arg/Lys-rich regions and motifs of Rep proteins, excluding putative hypervariable regions [26]. As shown in figure 5B, 4 groups (1–4) were created within one great cluster that differed from TLMV. Within this large cluster, YONBAN isolates formed a distinct outer group, indicating the first divergent species (group 1). Group 2 consisted of prototype TTV and group 3 was PMV. Finally, group 4, including the 8 SENV isolates, probably formed the last and most recent divergence from a common ancestor. Thus, prototype TTV, PMV, SENV (SANBAN and TUS01), and YONBAN probably evolved from a common ancestor.

Discussion

SENV belongs to the superfamily of TTV-related viruses. It is a small, most likely nonenveloped and single-stranded circular virus [3, 4]. Comparison of the genomic sequences, their encoded proteins, and biophysical characteristics of the virus may shed light on the evolutionary history of SENV in relation to TTV.

A phylogenetic analysis of 8 different SENV isolates, prototype TTV, and some TTV variants, including SANBAN, TUS01, PMV, and YONBAN, was performed by using near-complete ORF1 sequences, with TLMV as an outgroup. This analysis revealed 4 distinct groups; group 1, novel YONBAN isolates; group 2, the prototype TTV isolates; group 3, PMV only; and group 4, SENV, SANBAN, and TUS01 isolates. The 4 groups were created within one great cluster that differed from TLMV, suggesting that prototype TTV, PMV, SENV (SANBAN and TUS01), and YONBAN probably evolved from a common ancestor. Of note, group 1 apparently had diverged from a common ancestor earlier than the other groups. YONBAN may be the origin of TTV-related species. Group 4, including SENV, appeared to be the most recently divergent group. The genetic distances, based on the frequency distribution of pairwise genetic distances produced a range of 0.72145–0.93158 among the 4 groups comprising tier B (figure 4). The genetic distances of near-complete sequences among the 4 groups were greater than those between GBV-A [27] cloned from the serum of tamarin and GBV-C [28] identified from human serum samples (0.70357–0.73085). This suggests that the 4 groups are different virus species. However, such taxonomic matters should be determined by the International Committee on Taxonomy of Viruses or other special groups.

A large number of partial distinct sequences have been ob-

Figure 4. Pairwise genetic distances of entire open-reading frame (ORF) for 21 isolates, including SEN virus (SENV), prototype TT virus (TTV), and TTV variants. Genetic distances are on X-axis. Frequency of occurrence of discrete distances is on Y-axis. Tier A is composed of genetic distances of 0.33283–0.70492, which may be classified as genotypes; genetic distances among SEN isolates (0.40521–0.70492) were much greater than those among prototype TTV isolates (0.33283–0.38036). Tier B is composed of genetic distances of 0.72145–0.93158. Comparison of sequences for 8 novel SENV, prototype TTV, PMV, and YONBAN isolates used in this study show such large genetic distances distributed within tier B, which may be interpreted as different virus species. Genetic distances between SENV-C and SENV-H, SENV-D and SENV-F, and SENV-B and TUS01 are between 0.23348 and 0.29066 (*). Thus, these variants are intermediate between TTV subtypes and TTV genotypes.

Discussion

SENV belongs to the superfamily of TTV-related viruses. It is a small, most likely nonenveloped and single-stranded circular virus [3, 4]. Comparison of the genomic sequences, their encoded proteins, and biophysical characteristics of the virus may shed light on the evolutionary history of SENV in relation to TTV.

A phylogenetic analysis of 8 different SENV isolates, prototype TTV, and some TTV variants, including SANBAN, TUS01, PMV, and YONBAN, was performed by using near-complete ORF1 sequences, with TLMV as an outgroup. This analysis revealed 4 distinct groups; group 1, novel YONBAN isolates; group 2, the prototype TTV isolates; group 3, PMV only; and group 4, SENV, SANBAN, and TUS01 isolates. The 4 groups were created within one great cluster that differed from TLMV, suggesting that prototype TTV, PMV, SENV (SANBAN and TUS01), and YONBAN probably evolved from a common ancestor. Of note, group 1 apparently had diverged from a common ancestor earlier than the other groups. YONBAN may be the origin of TTV-related species. Group 4, including SENV, appeared to be the most recently divergent group. The genetic distances, based on the frequency distribution of pairwise genetic distances produced a range of 0.72145–0.93158 among the 4 groups comprising tier B (figure 4). The genetic distances of near-complete sequences among the 4 groups were greater than those between GBV-A [27] cloned from the serum of tamarin and GBV-C [28] identified from human serum samples (0.70357–0.73085). This suggests that the 4 groups are different virus species. However, such taxonomic matters should be determined by the International Committee on Taxonomy of Viruses or other special groups.

A large number of partial distinct sequences have been ob-
A constructed on basis of 21 isolates of entire open-reading frame (ORF) 1±3 (2612 nucleotides [nt]; all available codon positions). These 21 isolates comprise 8 SEN virus (SENV) isolates, 6 prototype TT virus (TTV) isolates (TA278, JA20, JA2B, JaCHC2c, JaBD89, and T3PB), and 7 TTV variants (SANBAN, TTVsan-IR1031, TUS01, TUPB, PMV, TTVyon-LC011, and TTVyon-KC186) obtained from DNA Data Bank of Japan, European Molecular Biology Laboratory, and GenBank nt sequence databases. Horizontal bar indicates no. of nt substitutions per site. This shows great diversity and 4 distinct clusters (groups 1±4) separated by a bootstrap value of 100%. Eight SENV isolates form a putative fourth cluster (group IV), including SANBAN and TUS01 isolates.

B, constructed on basis of available nt sequences (1485 nt; all codon positions) of ORF1 by using TTV-like minivirus (TLMV-NLC030) as an outgroup. Groups are as follows: 1, YONBAN isolates form distinct outer group; 2, prototype TTV isolates; 3, PMV; and 4, SENV, SANBAN, and TUS01 isolates form the last and most recent divergent species.

The prevalence of TTV-related viruses, including SENV, using universal primers was very high among blood donors in Japan (>90%) [31]. In addition, nonhuman primates were also infected with TTV-related viruses with marked sequence divergence from human TTV [32]. Chimpanzees were infected with human G1a TTV, confirming cross-species infection [22]. Some TTV isolates of chimpanzees were closer to certain genotypes of human TTV [32]. Our suggested classification will elucidate the evolution and relationship of each TTV-related species. This may involve combining human and nonhuman TTV-related species. Further evolutionary study may be needed when additional viruses are discovered.

SENV possessed a small circular ssDNA genome structurally similar to circoviruses but with poor sequence homology. In the N-terminus of ORF1, the highly basic Arg/Lys-rich domains with particularly high hydrophilicity, similar to those in the coat protein (VP1) of circoviruses [24], were conserved among the different SENV isolates except for SENV-A and SENV-B. The absence of this highly basic region of SENV-A...
and SENV-B might result in different biologic functions. Moreover, the Rep proteins involved in rolling circle replication possessed 4 motifs conserved among many plant and animal circoviruses [24]. Most SENV isolates also had 2 or 3 of the 4 conserved motifs and were predicted to replicate by a rolling circle mechanism, although we lacked evidence to suggest the function of this motif.

The deduced aa sequences encoded by putative ORF3 (pORF3) had an S-rich tract preceded by a cluster of basic aas (R and K). Previous studies have shown that the pORF3 has had an S-rich tract preceded by a cluster of basic aas and may play an important role in the replication of such ssDNA viruses [16]. Our BLASTP research indicates that pORF3, which contains some nuclear targeting sequences at the N-terminus followed by S-rich sequences, has several DNA- or RNA-binding nuclear proteins (figure 3). This homology leads to the speculation that the pORF3 may be a DNA- or RNA-binding nuclear protein mediating either transcriptional repression or activation.

In the SENV grouping, SENV-D and SENV-H are associated with transfusion-associated hepatitis [5], but causality has not been proved. Within the 4 proposed genomic groupings and certainly between groupings, the members of this family are so genomically diverse that some could be pathogenic and others without disease consequences. The extreme diversity dictates that the search for disease associations must be performed for each individual member of the family or for small groupings of the most closely related members. This search should not be restricted to cryptogenic liver disease and should also investigate associations with other diseases that are now of unknown etiology.

Figure 6.  Pairwise genetic distances of partial N22 region obtained from 55 isolates, including distinct groups. These are 8 SEN virus isolates, prototype TT virus (TTV) genotypes 1–6, and several other TTV variants, including SANBAN, YONBAN, TUS01, PMV, TTV genotypes 7–16 [10], TJN01, TJN02 [29], and TTV-2-TTV-5 reported by Khudyakov et al. [6]. Broad single tier of sequence diversity is revealed.

References


