Full-Length Sequences of One Genotype 4 and Three Genotype 3 Hepatitis E Viruses in Fecal Samples from Domestic Swine in Japan

Takeru Urayama1,2, Sompong Sapsutthipas1, Munet Tsujikawa2, Akifumi Yamashita3, Hiromi Nishigaki2, Madiha S. Ibrahim1,4, Katsuro Hagiwara5, Mikihiro Yunoki1,2, Teruo Yasunaga3, Teruhide Yamaguchi6 and Kazuyoshi Ikuta*,1

1Department of Virology, International Center for Infectious Disease Control, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan
2Infectious Pathogen Research Group, Osaka Laboratory, Benesis Corporation, Yodogawa-ku, Osaka 532-8505, Japan
3Department of Genome Informatics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan
4Department of Microbiology, Faculty of Veterinary Medicine, Damanhour branch, EL-Bostan, Alexandria University, Alexandria, Egypt
5Department of Veterinary Microbiology, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu, Hokkaido 069-8501, Japan
6National Institute of Health Sciences, Division of Biological Chemistry and Biologicals, Setagaya-ku, Tokyo 158-8501, Japan

Abstract: The Hepatitis E virus (HEV) induces zoonotic infections and causes hepatitis. In Japan, HEV occurs in deer, wild boar and swine, and genotype (G)3 and G4 have been isolated from domestic swine. We previously reported that HEV isolates from a total of 320 swine fecal samples from 32 farms in Japan could be predominantly classified into four clusters: three G3 (G3JP, G3SP and G3 US) and one G4 (G4JP). In this study, we performed full-length sequencing of four representative HEVs, one from each of the clusters. We found significant nucleotide variation throughout the sequences within a genotype, but not within each cluster. However, we found few variations at the amino acid level. Most of the highly conserved regions within genotypes were concentrated in the overlapping region of open reading frame (ORF)2 and ORF3, while most of the variable regions were within the ORF1 V region. Essentially, this region was highly conserved among G3 clusters, with some more dissimilarities between G3 SP and the other two clusters, G3 JP and G3 US. The regions conserved and variable across genotypes had virtually the same positions as those within genotypes, but were much narrower and wider, respectively. For the latter, ORF1 V and P regions were especially variable. Finally, we focused on the sequence conservation in the region widely used for primer and probe sets to detect HEV infections.

Keywords: HEV, full-length sequence, swine, feces, Japan.

INTRODUCTION

Several microbial agents induce hepatitis, but only HEV causes zoonotic hepatitis, mainly through food-borne transmission from domestic swine, wild boar and wild deer via the ingestion of uncooked or undercooked meat [1-5].

HEV is a non-enveloped small (27-34 nm in diameter) virus of the genus Hepeviridae (6). It contains single-stranded, positive-sense RNA of approximately 7.3 kilobases and its sequences have been classified into G1 to G4 [6, 7].

In Japan, a high prevalence of swine anti-HEV (71%) among swine aged 3-6 months and a high rate of viremia (11%) among swine of 2-4 months have been reported [8, 9], and two types of HEV are mainly circulating among pigs and humans. One type consists of three clusters of G3 (G3JP, G3SP and G3 US), and the other consists of one cluster of G4 (G4JP), which correspond to subgenotypes 3b, 3e, 3a and 4c, respectively [10]. In this study, HEVs representative of these four clusters isolated from swine feces were analyzed and their genomic and amino acid sequences compared.

MATERIALS AND METHODS

Sampling

A total of 320 fecal samples from 32 commercial pig farms (1 sample from each of the pig houses on individual farms) in Japan were processed for the partial purification of HEV as reported [11].
Amplification and Sequencing of Full-Length HEV Genomes

Total RNA was subjected to RT-PCR for amplification of the nearly full-length 7kb sequence of the HEV genome. The RNA was reverse-transcribed with PrimeScript Reverse Transcriptase (TaKaRa Biomedicals, Shiga, Japan) and subjected to a first round of amplification using PrimeSTAR GXL DNA Polymerase (TaKaRa Biomedicals). The primers used for reverse transcription, the first round of PCR, and any second round of PCR are indicated in Fig. (1). Nearly full-length cDNA of G3Jp, G3Sp and G3Us could be generated in the first round. That of G4Jp was generated in the second round. The 5’-end sequence was determined by an RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) technique with the First Choice RLM-RACE kit (Ambion, Austin, TX). RNA was treated with calf intestinal alkaline phosphatase followed by tobacco acid pyrophosphatase, and then ligated to an RNA adaptor supplied in the kit. RT-PCR was performed with an OneStep RT-PCR Kit (QIAGEN, Hilden, Germany). The second round of PCR was performed using Hot Start Taq DNA Polymerase (QIAGEN). The primers for these PCRs are also shown in Fig. (1). The 3’-end sequence was determined with a 3’-Full RACE Core Set (TaKaRa Biomedicals) according to the directions. Primers for 3’-end amplification are shown in Fig. (1). The 5’-terminal and 3’-terminal PCR products were ligated with pT7Blunt (Novagen, San Diego, CA) and cloned into E. coli DH5α (TOYOBO, Osaka, Japan).

The cDNAs of nearly full length were sequenced directly using the BigDye Terminator ver1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) on an ABI PRISM 3100-Avant Genetic Analyzer (Applied Biosystems). The cloned 5’-end and 3’-end cDNAs were also sequenced using this kit.

Sequence Analysis of PCR Products

The amplification products were extracted from agarose gels using a QIAEX II Gel Extraction Kit (QIAGEN). Both strands of the products were sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems). Sequence analysis was performed using Genetyx version 7.0 (Genetyx, Tokyo, Japan). Sequence alignments were generated by CLUSTAL W (version 1.8) [12]. A phylogenetic tree was constructed by the neighbor-joining method, as described previously [13], based on the entire nucleotide sequence of the HEV genome. Bootstrap values were determined on 1,000 re-samplings of the data sets [14]. The final tree was obtained using the Tree View program (version 1.6.6) [15].

Sequence Comparison

The individual sequences after alignments were compared at the nucleotide level using the “Average Difference of All pairwise comparisons” mode in the DifferencePlot (http://www.gen-info.osaka-u.ac.jp/~uhmin/study/differencePlot/index.html) with a window size of 30 and 6 slides. Similarly, amino acid sequences were compared using a window size of 1 and 1 slide. Briefly, the “Average Difference of all pairwise comparisons” calculates match rates of each window for all possible sequence pairs, and then calculates the average match rate of all the pairs for each window. Finally, a DifferencePlot displays the average match rate as two types of graph (A. Yamashita and T. Yasunaga, Personal Communication).

RESULTS

Full-Length Sequencing of Four Representative HEVs Derived from Swine Fecal Samples in Japan

A total of 320 fecal samples were collected from 32 commercial farms in Japan. Among them, the 159 HEV positive samples were subjected to genotyping by RT-PCR with primers at ORF2 followed by sequencing [11]. All the HEV sequences, except for several unclassified types, were classified into four clusters, G3Jp, G3Sp, G3Us and G4Jp. Therefore, we selected a representative sample from each of the clusters: swJR-P5 (G3Jp), swJB-E10 (G3Sp), swJB-M8 (G3Us) and swJB-H7 (G4Jp).

The original fecal samples used for genotyping were also used for partial purification of these four HEVs. Extracted nucleic acid sequences were used to sequence the products of RT-PCR (Fig. 1). The full-length sequences showed heterogeneous populations, i.e., swJB-E10 having nucleotide variation at 17 sites; swJR-P5 and swJB-H7 having nucleotide variation at 1 site; and swJB-M8 with no variation (data not shown). Therefore, we selected the predominant nucleotide sequences. Sequence accession numbers are shown in Fig. (2).

Comparison of Full-Length Sequences within Each Genotype and Cluster

Full-length genomic sequences of HEVs belonging to G1 (n=5), G2 (n=1), G3 (n=23) and G4 (n=13) that were available in the GenBank database were used for the comparison. Consistent with our previous analysis using sequences for part of ORF2 [11], a phylogenetic tree of the full-length genomic sequences showed that our HEV samples are located within the four individual clusters, G3Jp, G3Sp, G3Us and G4Jp (Fig. 2).

Next, we estimated the variation at the nucleotide and amino acid levels, with all the sequences used for the phylogenetic analysis (Fig. 3A). The variation was estimated by comparing sequences every 30 nucleotides window as well as every amino acid. The regions highly conserved (>90% identity) among G3 and G4 at the nucleotide level were located around ORF3. This result indicates that although the HEV sequences belonging to G3 formed independent sub-clusters, significant parts remained highly conserved at the nucleotide level. In contrast, the variable regions (~60% and 60-74% identity) were concentrated in the ORF1 V region in both genotypes. Although most of the regions were conserved at the amino acid level, the V region of ORF1 was highly variable in both G3 and G4. Further, the variation within clusters at the nucleotide level was examined using individual HEV isolates based on the full-length sequences used for Fig. (2) (Fig. 3B). The results clearly showed for all four clusters that although the V region of ORF1 was variable, all the other regions were fairly well conserved. Comparisons among individual clusters revealed that the highly conserved regions spread over the full-length sequence were smaller in G3Jp. Thus, the individual isolates in a cluster showed some variation.
Fig. (1). Location and direction of the primers used for the sequence analysis of full-length HEV genomes. A) The primers for RT-PCR amplify nearly all of G3JP, G3SP and G3US. B) G4p was subjected to nested RT-PCR. For the 5'- and 3'-ends, the analysis was performed as described in Materials and Methods. C) Nucleotide sequences of primers and nucleotide positions are shown. The numbers in parentheses are the nucleotide positions of swJ570 (Accession No. 073912).
Comparison of Full-Length Sequences Across Clusters and Genotypes

Next, we tried a similar approach across the three clusters within G3 at the nucleotide level (Fig. 4A). Comparisons of swJB-P5 (G3JP) vs swJB-E10 (G3SP), swJB-E10 (G3SP) vs swJB-M8 (G3US), and swJB-P5 (G3JP) vs swJB-M8 (G3US), revealed a clear distribution of variable and conserved regions over the entire sequence. Interestingly, highly conserved regions (identity in >90%) were more frequent between G3JP and G3US than the other combinations. In contrast, variable regions (<60% and 60-74% identity) were identified in the V region of ORF1, but significantly smaller between G3JP and G3US. These results indicate that G3JP and G3US are more closely related than the other strains.

Further, we applied a similar approach to the cross-genotype variation between G3 and G4 (Fig. 4B). As expected, positions of highly conserved and variable regions were consistent with those of the previous combinations shown above. However, for all three combinations [swJB-H7 (G4JP) vs swJB-P5 (G3JP), swJB-H7 (G4JP) vs swJB-E10 (G3SP), and swJB-H7 (G4JP) vs swJB-M8 (G3US)], the highly conserved regions (identity in >90%) were significantly narrower, whereas the variable regions were apparently wider, throughout the genomic sequence, compared with the above comparisons within genotypes (Fig. 3A), within clusters (Fig. 3B), and between clusters (Fig. 4A).

As summarized in Table 1, comparison of the full-length sequences showed similar results, i.e., G3JP is very similar to G3US, somewhat different to G3SP, and very different to G4JP at the nucleotide level. This tendency was also observed at the amino acid level in ORF1 and ORF3. However, for ORF2, G3JP was very similar to G4JP as well as G3US and G3SP.

Variation in the Sequences for the Primers and Probes Used for Detection of the HEV Genome by RT-PCR

The comparison among the four clusters in G3 and G4 showed that the overlapping region of ORF2 and ORF3 was
A. Within genotypes

23 isolates of G3

Nucleotide

Amino acid

13 isolates of G4

Nucleotide

Amino acid

B. Within clusters

10 isolates of G3_{JP}

Nucleotide

4 isolates of G3_{SP}

Nucleotide

6 isolates of G3_{US}

Nucleotide

7 isolates of G4_{JP}

Nucleotide

Fig. (3). Conservation within genotypes and clusters. A) A total of 23 isolates belonging to G3, including three HEVs from this study, and a total of 13 isolates belonging to G4 including one HEV from this study, were used. For the sequence comparison, the average rate of conservation (%) every 30 nucleotides or every amino acid was calculated and values are shown with different colors. B) Ten isolates including swJR-P5 in the G3_{JP} cluster, 4 isolates including swJB-E10 in the G3_{SP} cluster, 6 isolates including swJB-M8 in the G3_{US} cluster and 7 isolates including swJB-H7 in the G4_{JP} cluster were used.
Fig. (4). Conservation across clusters in G3 and between G3 and G4 at the nucleic acid level. Two of the four isolates in this study, G3<sub>JP</sub>, G3<sub>SP</sub>, G3<sub>US</sub> and G4<sub>JP</sub>, were compared. Sequences with different conservation rates are shown in different colors, as in Fig. (3).
amplification in the overlapping region of ORF2 and ORF3 focused on the appropriateness for this region as a target for detecting HEV RNA.

An example of a table from the text:

<table>
<thead>
<tr>
<th>% Identity in Nucleotides; Full-Length</th>
<th>swJR-P5 (G3JP)</th>
<th>swJB-M8 (G3SP)</th>
<th>swJB-E10 (G3SP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>swJB-M8 (G3SP)</td>
<td>87.4</td>
<td>81.7</td>
<td>75.6</td>
</tr>
<tr>
<td>swJB-E10 (G3SP)</td>
<td>81.7</td>
<td>81.3</td>
<td>77.4</td>
</tr>
<tr>
<td>swJB-H7 (G4SP)</td>
<td>75.6</td>
<td>77.4</td>
<td>75.5</td>
</tr>
</tbody>
</table>

% Identity in Amino Acids; ORF1

<table>
<thead>
<tr>
<th>% Identity in Amino Acids; ORF1</th>
<th>swJR-P5 (G3JP)</th>
<th>swJB-M8 (G3SP)</th>
<th>swJB-E10 (G3SP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>swJB-M8 (G3SP)</td>
<td>96.8</td>
<td>93.9</td>
<td>85.7</td>
</tr>
<tr>
<td>swJB-E10 (G3SP)</td>
<td>93.9</td>
<td>93.6</td>
<td>85.3</td>
</tr>
<tr>
<td>swJB-H7 (G4SP)</td>
<td>85.7</td>
<td>85.3</td>
<td>85.7</td>
</tr>
</tbody>
</table>

% Identity in Amino Acids; ORF2

<table>
<thead>
<tr>
<th>% Identity in Amino Acids; ORF2</th>
<th>swJR-P5 (G3JP)</th>
<th>swJB-M8 (G3SP)</th>
<th>swJB-E10 (G3SP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>swJB-M8 (G3SP)</td>
<td>98.5</td>
<td>97.4</td>
<td>91.8</td>
</tr>
<tr>
<td>swJB-E10 (G3SP)</td>
<td>97.4</td>
<td>97.3</td>
<td>91.5</td>
</tr>
<tr>
<td>swJB-H7 (G4SP)</td>
<td>91.8</td>
<td>91.5</td>
<td>92.1</td>
</tr>
</tbody>
</table>

% Identity in Amino Acids; ORF3

<table>
<thead>
<tr>
<th>% Identity in Amino Acids; ORF3</th>
<th>swJR-P5 (G3JP)</th>
<th>swJB-M8 (G3SP)</th>
<th>swJB-E10 (G3SP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>swJB-M8 (G3SP)</td>
<td>96.5</td>
<td>93.8</td>
<td>82.5</td>
</tr>
<tr>
<td>swJB-E10 (G3SP)</td>
<td>93.8</td>
<td>92.0</td>
<td>83.3</td>
</tr>
<tr>
<td>swJB-H7 (G4SP)</td>
<td>82.5</td>
<td>83.3</td>
<td>82.5</td>
</tr>
</tbody>
</table>

The transmission of HEV to humans in Japan is mostly food-borne [1-5]. Full-length sequences of four HEVs in fecal samples from pig farms in Japan were analyzed in this study. The four representatives were located in independent clusters in a phylogenetic tree constructed with these sequences together with full-length reference sequences from the GenBank database: G3JP, G3SP, G3US, and G4JP covered most of the HEVs distributed among swine as well as humans in Japan. A comparison of these four sequences with the reference sequences revealed highly conserved regions that overlap between ORF2 and ORF3, as well as variable regions in the ORF1 V region. Positions of the variable as well as conserved regions were virtually conserved across clusters and genotypes, as well as within clusters and genotypes.

Comparisons between clusters revealed higher genetic similarity between G3JP and G3US, which were slightly more distant from G3SP. We also found that the regions highly conserved among G3 and G4 occur within smaller areas in the overlapping region of ORF2 and ORF3. In addition, not only V but also P regions of ORF1 were highly variable. Interestingly, the ORF1 P region is unlikely to code for a papain-like protease [20], because only a part of the HEV sequence can encode the protease (see the sequences listed by their accession numbers in Fig. 2).

Finally, we examined how efficiently the primer and probe sets could amplify HEV RNA. The G3SP and G3US HEVs arrived in Japan through the importation of pigs from the United Kingdom [21] and United States [8], respectively. G3SP and G3US, in addition to G3JP and G4JP, are currently distributed in pigs and humans in Japan. These HEV clusters were highly conserved at the amino acid level, but highly variable at the nucleotide level. Therefore, a sensitive method of detecting the genome is urgently needed. Several groups have used primer and probe sets at similar sites [17-19]. We quantified the amount of HEV RNA according to the method of Jothikumar et al. [17]. Sequences from the GenBank database were used for the examination: 72 sequences from animals including pigs and 107 sequences from humans at HE86, as shown in Table 2. HE86, FHE88 and HE87 correspond to JVEHVF, JVEHVP and JVEHVR, respectively, in Jothikumar et al. [17]. Some 96.3%, 94.4% and 96.3% of the human-derived HEV sequences were conserved in HE86, FHE88 and HE87, respectively. Slightly less conservation, except for the region at HE87, was observed in the sequences from animals: 91.7%, 88.9% and 97.2% in HE86, FHE88 and HE87, respectively.

### ACKNOWLEDGEMENTS

We thank Dr. Hiroshi Yasue, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan for valuable discussions. This work was conducted based on collaborative research projects between Osaka University and Benesis Corporation and between Rakuno Gakuen University and Benesis Corporation, and supported in part by a grant from the Japan Human Sciences Foundation (Grant number KHB1011).
### Table 2. Nucleotide Variation in Regions Used for Primer and Probe Sets to Detect the HEV Genome with High Sensitivity

#### Sense Primer

<table>
<thead>
<tr>
<th>Virus Origin</th>
<th>Nucleotide Position 5286–5303 of swJ570 (AB073912)</th>
<th>Sequence Number</th>
<th>Appearance Ratio (%)</th>
<th>Primer Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swine, Wild boar, Deer and others, including of unknown origin</td>
<td>GGT GGT TTC TGG GGT GAC</td>
<td>66</td>
<td>91.7 %</td>
<td>HE86 (JVHEVF)</td>
</tr>
<tr>
<td>Human</td>
<td>GGT GGT TTC TGG GGT GAC</td>
<td>103</td>
<td>96.3 %</td>
<td>HE86 (JVHEVF)</td>
</tr>
</tbody>
</table>

#### Probe

<table>
<thead>
<tr>
<th>Virus Origin</th>
<th>Nucleotide Position 5309–5326 of swJ570 (AB073912)</th>
<th>Sequence Number</th>
<th>Appearance Ratio (%)</th>
<th>Probe Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swine, Wild boar, Deer and others, including of unknown origin</td>
<td>TGA TTC TCA GCC CTT CGC</td>
<td>64</td>
<td>88.9 %</td>
<td>FHE88 (JVHEVP)</td>
</tr>
<tr>
<td>Human</td>
<td>TGA TTC TCA GCC CTT CGC</td>
<td>101</td>
<td>94.4 %</td>
<td>FHE88 (JVHEVP)</td>
</tr>
</tbody>
</table>

#### Anti-Sense Primer

<table>
<thead>
<tr>
<th>Virus Origin</th>
<th>Nucleotide Position 5338–5355 of swJ570 (AB073912)</th>
<th>Sequence Number</th>
<th>Appearance Ratio (%)</th>
<th>Primer Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swine, Wild boar, Deer and others, including of unknown origin</td>
<td>TTC ATC CAA CCA ACC CCT</td>
<td>70</td>
<td>97.2 %</td>
<td>HE87 (JVHEVR)</td>
</tr>
<tr>
<td>Human</td>
<td>TTC ATC CAA CCA ACC CCT</td>
<td>103</td>
<td>96.3 %</td>
<td>HE87 (JVHEVR)</td>
</tr>
</tbody>
</table>

Probe FHE100, differing by one nucleotide from FHE88, was designed to detect swJB-M8 (G3x3).

### REFERENCES


Full-Length Sequences of One Genotype 4 and Three Genotype 3 HEV


