

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 G3_{JP}, G3_{SP} and G3_{US} could be generated in the first round. That of G4_{JP} 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 pT7Blue (Novagen, San Diego, CA) and cloned into *E. coli* DH5 α (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, G3_{JP}, G3_{SP}, G3_{US} and G4_{JP}. Therefore, we selected a representative sample from each of the clusters: swJR-P5 (G3_{JP}), swJB-E10 (G3_{SP}), swJB-M8 (G3_{US}) and swJB-H7 (G4_{JP}).

The original fecal samples used for genotyping were also used for partial purification of these four HEVs. Extracted nucleic acid samples 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, G3_{JP}, G3_{SP}, G3_{US} and G4_{JP} (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 G3_{JP}. Thus, the individual isolates in a cluster showed some variation.

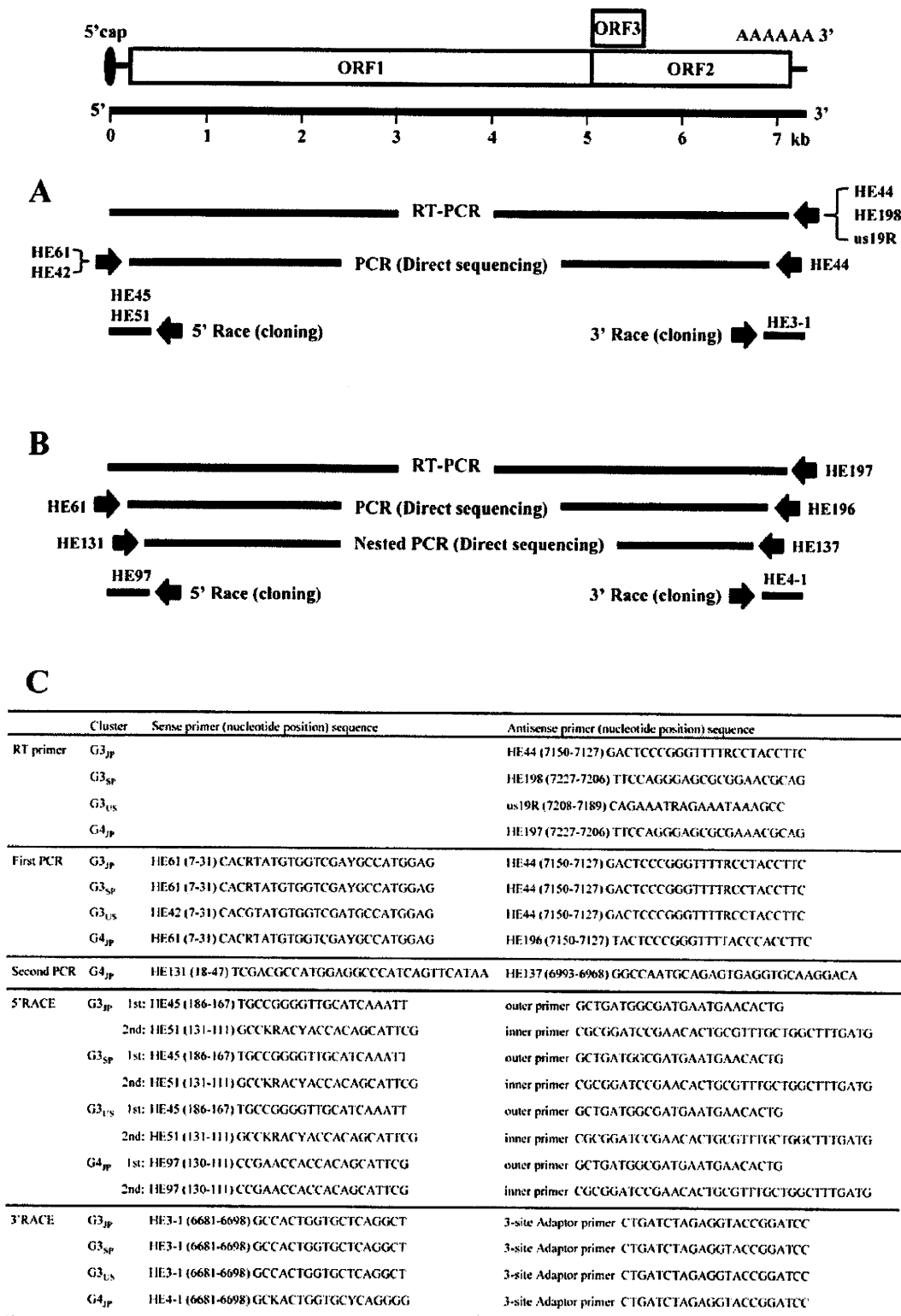


Fig. (1). Location and direction of the primers used for the sequence analysis of full-length HEV genomes. A) The primers for RT-PCR amplify nearly all of G3_{JP}, G3_{SP} and G3_{US}. B) G4_{JP} was subjected to nested RT-PCR. For the 5'- and 3'-ends, the analysis was performed as described in Materials 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).

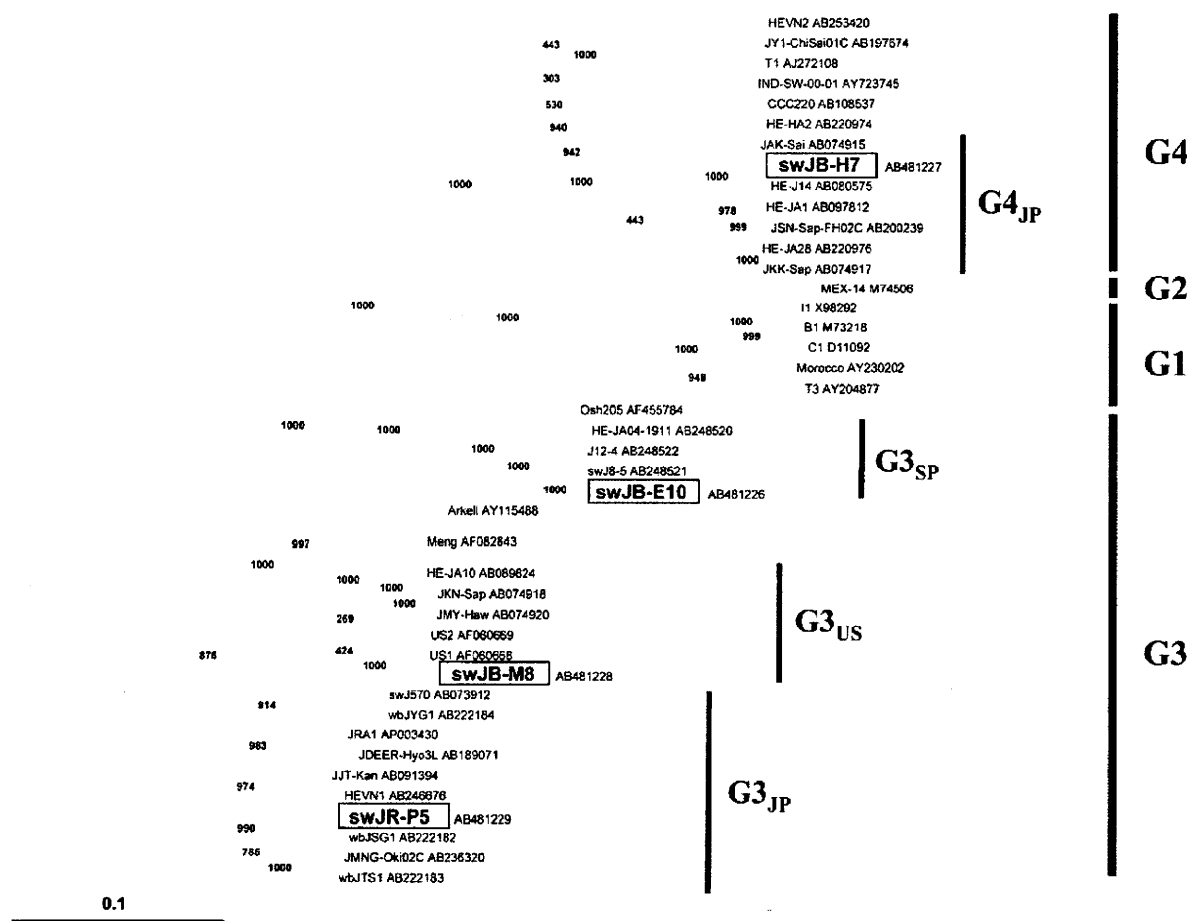


Fig. (2). Phylogenetic tree with full-length sequences of the four HEVs, together with reference HEVs. The name and accession number of individual isolates belonging to G1 to G4 are shown. The four representative HEVs in this study are boxed. The number on each branch is a bootstrap value from 1,000 re-samplings.

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 (G3_{JP}) vs swJB-E10 (G3_{SP}), swJB-E10 (G3_{SP}) vs swJB-M8 (G3_{US}), and swJB-P5 (G3_{JP}) vs swJB-M8 (G3_{US}), revealed a clear distribution of variable and conserved regions over the entire sequence. Interestingly, highly conserved regions (identity in >90%) were more frequent between G3_{JP} and G3_{US} 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 G3_{JP} and G3_{US}. These results indicate that G3_{JP} and G3_{US} 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

(G4_{JP}) vs swJB-P5 (G3_{JP}), swJB-H7 (G4_{JP}) vs swJB-E10 (G3_{SP}), and swJB-H7 (G4_{JP}) vs swJB-M8 (G3_{US})], 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., G3_{JP} is very similar to G3_{US}, somewhat different to G3_{SP}, and very different to G4_{JP} at the nucleotide level. This tendency was also observed at the amino acid level in ORF1 and ORF3. However, for ORF2, G3_{JP} was very similar to G4_{JP} as well as G3_{US} and G3_{SP}.

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

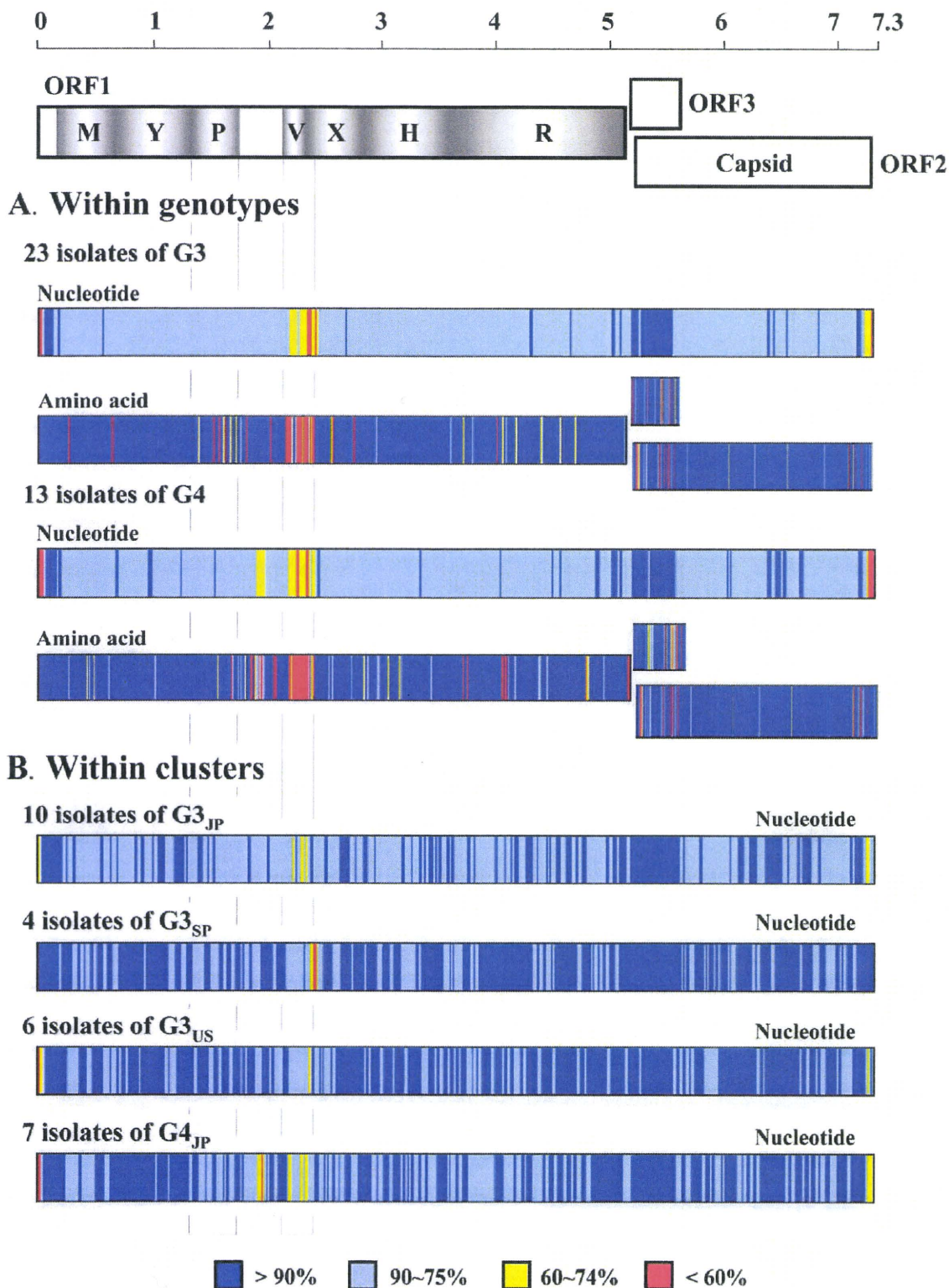
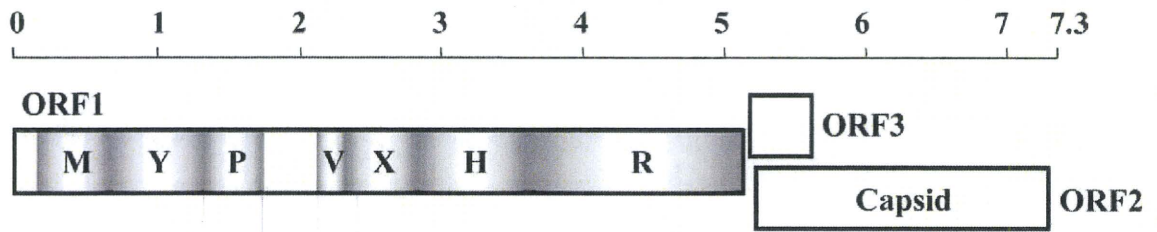
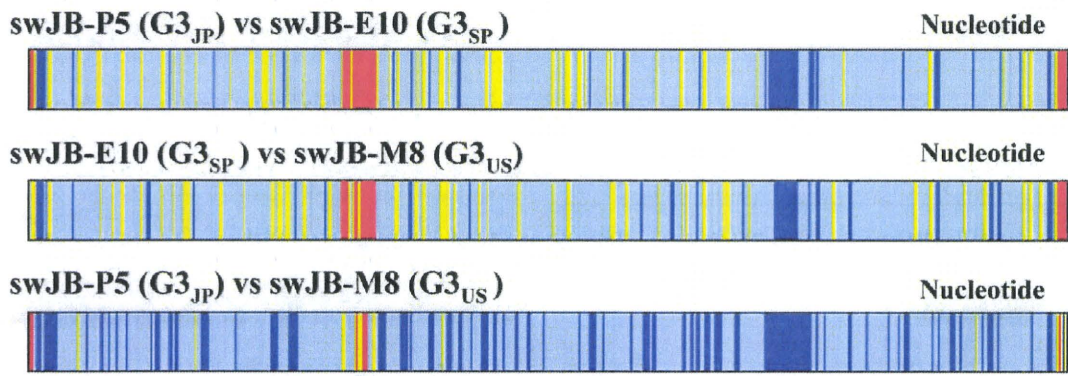


Fig. (3). Conservation within genotypes and clusters. A) A total of 23 isolates belonging to G3, including three HEVs from this study, and a total of 13 isolates belonging to G4 including one HEV from this study, were used. For the sequence comparison, the average rate of conservation (%) every 30 nucleotides 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.



A. Between clusters in G3



B. Between genotypes

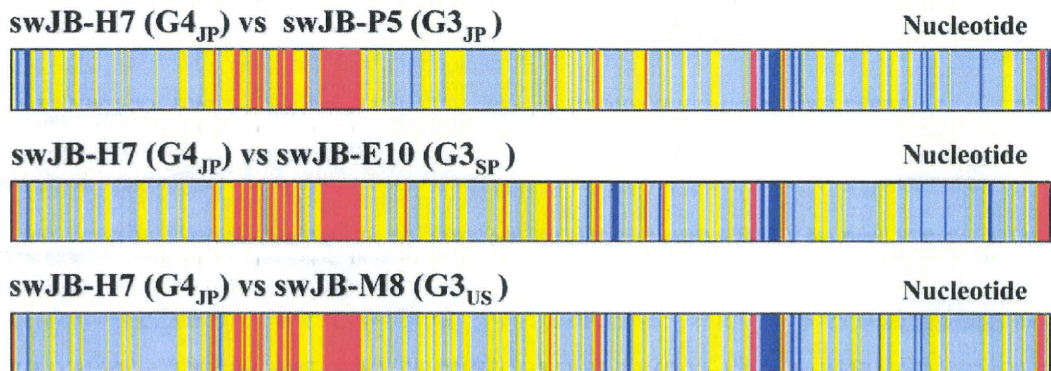


Fig. (4). Conservation across clusters in G3 and between G3 and G4 at the nucleic acid level. Two of the four isolates in this study, G3_{JP}, G3_{SP}, G3_{US} and G4_{JP}, were compared. Sequences with different conservation rates are shown in different colors, as in Fig. (3).

Table 1. Homology of Individual ORF Sequences of Four HEV Isolates**% Identity in Nucleotides; Full-Length**

	swJR-P5 (G3 _{JP})	swJB-M8 (G3 _{US})	swJB-E10 (G3 _{SP})
swJB-M8 (G3 _{US})	87.4		
swJB-E10 (G3 _{SP})	81.7	81.3	
swJB-H7 (G4 _{JP})	75.6	77.4	75.5

% Identity in Amino Acids; ORF1

	swJR-P5 (G3 _{JP})	swJB-M8 (G3 _{US})	swJB-E10 (G3 _{SP})
swJB-M8 (G3 _{US})	96.8		
swJB-E10 (G3 _{SP})	93.9	93.6	
swJB-H7 (G4 _{JP})	85.7	85.3	85.7

% Identity in Amino Acids; ORF2

	swJR-P5 (G3 _{JP})	swJB-M8 (G3 _{US})	swJB-E10 (G3 _{SP})
swJB-M8 (G3 _{US})	98.5		
swJB-E10 (G3 _{SP})	97.4	97.3	
swJB-H7 (G4 _{JP})	91.8	91.5	92.1

% Identity in Amino Acids; ORF3

	swJR-P5 (G3 _{JP})	swJB-M8 (G3 _{US})	swJB-E10 (G3 _{SP})
swJB-M8 (G3 _{US})	96.5		
swJB-E10 (G3 _{SP})	93.8	92.0	
swJB-H7 (G4 _{JP})	82.5	83.3	82.5

highly conserved. This makes the region a potential target for detecting HEV RNA by RT-PCR. Therefore, we next focused on the appropriateness for this region as a target for detecting HEV RNA. Several groups have performed amplification in the overlapping region of ORF2 and ORF3 [16-19]. 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 and HE87; and 72 sequences from animals including pigs and the same 107 sequences from humans at FHE88, as shown in Table 2. HE86, FHE88 and HE87 correspond to JVHEVF, JVHEVP and JVHEVR, 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.

DISCUSSION

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: G3_{JP}, G3_{SP}, G3_{US} and G4_{JP} 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 G3_{JP} and G3_{US}, which were slightly more distant from G3_{SP}. 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 G3_{SP} and G3_{US} HEVs arrived in Japan through the importation of pigs from the United Kingdom [21] and United States [8], respectively. G3_{SP} and G3_{US}, in addition to G3_{JP} and G4_{JP}, 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]. The results suggest that most of the sequences from GenBank (>88%) that were derived from humans and animals would be amplified with the primers HE86 (JVHEVF)/HE87 (JVHEVP) and probe FHE88 (JVHEVR). G3_{US} differs by one nucleotide from the other three isolates in the region of the probe. The sequence of probe FHE100 for G3_{US} was found in only 0.9% and 2.8% of the sequences from humans and animals, respectively. Based on our experience with the detection of G3_{US}, FHE100 was necessary for the detection of G3_{US}, because FHE88 was more than 10-times less sensitive than FHE100. Thus, most of the heterogeneous genomic sequences among HEVs in humans in Japan would be covered by these primer and probe sets, although this needs to be confirmed with experimental analyses. These four HEVs could be useful as a standard for heterogeneous HEV genomic sequences.

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).

- [6] Emerson SU, Anderson D, Arankalle A, *et al.* In: Fauquet CM, Ed. Hepatitis E virus. In: Virus taxonomy: eighth report of the international committee on taxonomy of viruses. Elsevier/Academic Press, London, 2004; 851-855.
- [7] Schlauder GG, Mushahwar IK. Genetic heterogeneity of hepatitis E virus. *J Med Virol* 2001; 65: 282-92.
- [8] Takahashi M, Nishizawa T, Miyajima H, *et al.* Swine hepatitis E virus strains in Japan form four phylogenetic clusters comparable with those of Japanese isolates of human hepatitis E virus. *J Gen Virol* 2003; 84: 851-62.
- [9] Takahashi M, Nishizawa T, Tanaka T, *et al.* Correlation between positivity for immunoglobulin A antibodies and viraemia of swine hepatitis E virus observed among farm pigs in Japan. *J Gen Virol* 2005; 86: 1807-13.
- [10] Lu L, Li C, Hagedorn CH. Phylogenetic analysis of global hepatitis E virus sequences: genetic diversity, subtypes and zoonosis. *Rev Med Virol* 2006; 16: 5-36.
- [11] Sapsutthipas S, Urayama T, Yamate M, *et al.* Sequence variation in hepatitis E virus genotypes 3 and 4 from swine fecal samples in Japan. *Open Vet Sci J* 2009; 3: 68-75.
- [12] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; 22: 4673-80.
- [13] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; 4: 406-25.
- [14] Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 1985; 39: 783-91.
- [15] Page RD. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 1996; 12: 357-8.
- [16] Enouf V, Dos Reis G, Guthmann JP, *et al.* Validation of single real-time TaqMan PCR assay for the detection and quantitation of four major genotypes of hepatitis E virus in clinical specimens. *J Med Virol* 2006; 78: 1076-82.
- [17] Jothikumar N, Cromeans TL, Robertson BH, *et al.* A broadly reactive one-step real-time RT-PCR assay for rapid and sensitive detection of hepatitis E virus. *J Virol Methods* 2006; 131: 65-71.
- [18] Orrù G, Masia G, Orrù G, *et al.* Detection and quantitation of hepatitis E virus in human faeces by real-time quantitative PCR. *J Virol Methods* 2004; 118: 77-82.
- [19] Zhao C, Li Z, Yan B, *et al.* Comparison of real-time fluorescent RT-PCR and conventional RT-PCR for the detection of hepatitis E virus genotypes prevalent in China. *J Med Virol* 2007; 79: 1966-73.
- [20] Koonin EV, Gorbalenya AE, Purdy MA, *et al.* Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proc Natl Acad Sci USA* 1992; 89: 8259-63.
- [21] Inoue J, Takahashi M, Ito K, *et al.* Analysis of human and swine hepatitis E virus (HEV) isolates of genotype 3 in Japan that are only 81-83 % similar to reported HEV isolates of the same genotype over the entire genome. *J Gen Virol* 2006; 87: 2363-9.

Received: November 24, 2009

Revised: February 25, 2010

Accepted: April 28, 2010

© Urayama *et al.*; licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.

