Analysis of the full-length genome of a hepatitis E virus isolate obtained from a wild boar in Japan that is classifiable into a novel genotype

Masaharu Takahashi,1,† Tsutomu Nishizawa,1,† Hiroyuki Sato,2,‡ Yukihiro Sato,3 Jirintai,1 Shigeo Nagashima1 and Hiroaki Okamoto1

1Division of Virology, Department of Infection and Immunity, Jichi Medical University School of Medicine, Shimotsuke-Shi, Tochigi-Ken 329-0498, Japan
2Department of Internal Medicine, Maniwa Kokuho Onsen Hospital, Maniwa-Shi, Okayama-Ken 717-0403, Japan
3Department of Internal Medicine, Kamiichi General Hospital, Nakanikawa-Gun, Toyama-Ken 930-0391, Japan

While performing a nationwide survey of hepatitis E virus (HEV) infection among 450 wild boars (Sus scrofa leucomystax) that had been captured in Japan between November 2005 and March 2010, we found 16 boars (3.6 %) with ongoing HEV infection: 11 had genotype 3 HEV, four had genotype 4 HEV and the remaining boar was infected with HEV of an unrecognized genotype (designated wbJOY_06). The entire wbJOY_06 genome was sequenced and was found to comprise 7246 nt excluding the poly(A) tail. The wbJOY_06 isolate was highly divergent from known genotype 1–4 HEV isolates derived from humans, swine, wild boars, deer, mongoose and rabbits (n=145) by 22.6–27.7 %, rat HEV isolates (n=2) by 46.0–46.2 %, and avian HEV isolates (n=5) by 52.5–53.1 % over the entire genome. A Simplot analysis revealed no significant recombination between the existing HEV strains of genotypes 1–4. Therefore, we propose that the wbJOY_06 isolate is the first member of a previously unidentified genotype.

The hepatitis E virus (HEV) was first identified as a leading cause of acute and fulminant hepatitis linked to faecal–oral transmission in tropical and subtropical countries. However, hepatitis E has been found to be endemic in industrialized countries, including Japan, the USA and European countries, where autochthonous HEV infections are an emerging concern (Dalton et al., 2008; Okamoto et al., 2003; Purcell & Emerson, 2008). It has recently been reported that zoonotic food-borne transmission of HEV from domestic pigs, wild boars and wild deer to humans plays an important role in the occurrence of cryptic hepatitis E in industrialized countries, including Japan and France, where people have distinctive habits of eating uncooked or undercooked meat (including the liver and colon/intestine of animals) (Colson et al., 2010; Matsuda et al., 2003; Tamada et al., 2004; Tei et al., 2003; Yazaki et al., 2003).

HEV is a non-enveloped virus and its genome is a positive-sense ssRNA, which is capped and polyadenylated (Kabrane-Lazizi et al., 1999; Tam et al., 1991). It is classified as the sole member of the genus Hepevirus in the family Hepeviridae (Emerson et al., 2005). The genome is approximately 7.2 kb in size and contains three ORFs that encode non-structural proteins involved in replication (ORF1), a capsid protein consisting of 660 aa (ORF2), and a small protein of only 113–114 aa (ORF3) that is essential for viral infectivity in animals (Graff et al., 2005; Huang et al., 2007) and virion egress (Emerson et al., 2010; Yamada et al., 2009a). Four genotypes of HEV that infect humans have been identified (Emerson et al., 2005; Lu et al., 2006; Okamoto, 2007). HEV genotypes 1 and 2 are restricted to humans and associated with outbreaks of hepatitis E as water-borne epidemics in developing countries, whereas HEV genotypes 3 and 4 are zoonotic and responsible for sporadic cases of hepatitis E worldwide.

Recently, significant progress has been made in understanding the animal reservoirs of HEV (Meng, 2010; Pavio et al., 2010). The discoveries of animal strains of HEV from domestic pigs (Meng et al., 1997), wild boars (Sonoda et al., 2004), deer (Tei et al., 2003), mongoose (Nakamura et al., 2006), chickens (Haqshenas et al., 2001; Payne et al., 1999), rabbits (Zhao et al., 2009) and rats (John et al., 2010) have...
significantly broadened the host range and genomic diversity of HEV. Regarding HEV from wild boars, HEV RNA and antibodies have been detected in several countries, including Australia, Germany, Hungary, Italy, Japan, the Netherlands and Spain, with a seroprevalence rate of 9–43% and the HEV RNA detection rate of 2–25% (Chandler et al., 1999; de Deus et al., 2008; Kaci et al., 2008; Martelli et al., 2008; Michitaka et al., 2007; Nishizawa et al., 2005; Rutjes et al., 2010; Sonoda et al., 2004). However, the genomic characteristics of boar HEV isolates are not fully understood. While performing a nationwide survey of HEV infection among 450 wild boars (Sus scrofa leucomystax) that had been captured in Japan between November 2005 and March 2010, we found 16 boars (3.6%) with ongoing HEV infection: 11 had genotype 3 HEV, four had genotype 4 HEV and the remaining boar was infected with HEV (designated wbJOY_06) that differed from reported HEV isolates by 18.0–24.3% within a 412 nt sequence of the ORF2 region, suggesting that wbJOY_06 is classifiable into an unrecognized genotype. Therefore, the present study was conducted to determine the full-length genomic sequence of the wbJOY_06 isolate and to clarify the genomic characteristics of the novel boar HEV isolate.

The wbJOY_06 isolate was recovered from liver tissues of a 20 kg male wild boar that had been caught in the forest of Maniwa city, Okayama Prefecture located in the western part of Honshu Island on 12 January 2006. To determine the full-length sequence of the wbJOY_06 genome, total RNA was extracted from boar liver specimens (500 mg) using the Trizol reagent (Invitrogen), and the RNA preparation thus obtained was reverse transcribed and subjected to nested PCR. The central 7 kb sequence of the wbJOY_06 genome was divided into 12 overlapping sections and amplified: they were nt 37–448 (412 nt) (primer sequences excluded), nt 386–1096 (711 nt), nt 1078–1270 (193 nt), nt 1247–2002 (756 nt), nt 1925–3232 (1308 nt), nt 3208–3912 (705 nt), nt 3895–4520 (626 nt), nt 4468–4624 (157 nt), nt 4613–5364 (752 nt), nt 5349–6013 (665 nt), nt 5991–6402 (412 nt) and nt 6388–7155 (768 nt). The extreme 5’-end sequence (nt 1–60) was determined by a modified rapid amplification of cDNA ends (RACE) technique called RNA ligase-mediated RACE (RLM-RACE) with the First Choice RLM-RACE kit (Ambion), as described previously (Okamoto et al., 2001). Amplification of the extreme 3’-end sequence [nt 7123–7246 excluding the poly(A) tail] was accomplished using the RACE method, according to a method described previously (Okamoto et al., 2001). The amplification products were sequenced on both strands either directly or after cloning into pT7BlueT-Vector (Novagen), and sequence analysis was performed as described previously (Okamoto et al., 2001).

The wbJOY_06 isolate had a genomic length of 7246 nt, excluding the poly(A) tract at the 3’ terminus, and possessed three major ORFs, similar to reported mammalian and avian HEV isolates (Emerson et al., 2005): ORF1, ORF2 and ORF3 code for 1709 aa (nt 26–5152), 660 aa (nt 5194–7173) and 112 aa (nt 5186–5521), respectively. The 5’ and 3’ untranslated regions of wbJOY_06 comprised 25 and 73 nt [excluding the poly(A) tail], respectively. Upon comparison with the 145 reported HEV genomes of genotypes 1–4 whose entire or nearly entire nucleotide sequences are already known, the wbJOY_06 genome shared nucleotide sequence identities of only 73.2–74.3% with human genotype 1 HEV (n = 20), 72.9% with human genotype 2 HEV (n = 1), 72.3–74.8% with genotype 3 HEV of human, swine, wild boar, deer, mongoose and rabbit origin (n = 66), and 76.3–77.4% with genotype 4 HEV of human, swine and wild boar origin (n = 58). Although rabbit HEV sequences were reported to belong to a novel genotype (Zhao et al., 2009), they segregated into genotype 3 in the present phylogenetic analyses (see Figs 1 and 2), corroborating a recent review article by Pavio et al. (2010).

In addition, wbJOY_06 was only 53.8–54.0% identical to rat HEV (n = 2) and 46.9–47.5% similar to avian HEV (n = 5) over the entire genome (Supplementary Table S1, available in JGV Online). These results indicate that the wbJOY_06 isolate is distantly related to the known HEV isolates of genotypes 1–4, including eight boar isolates of genotypes 3 and 4, and is clearly distinct from the previously reported rat and avian HEV isolates. The phylogenetic tree was constructed by the neighbour-joining method (Saitou & Nei, 1987), with the Kimura two-parameter correction model using online tools in the DDBJ database (National Institute of Genetics, Mishima, Japan), based on the overlapping (nearly) entire genomic sequence of the 146 HEV isolates [reported isolates (nos. 001–145: see Supplementary Table S2, available in JGV Online) and wbJOY_06], and bootstrap values were determined on 1000 resamplings of the datasets (Felsenstein, 1985). The tree confirmed that wbJOY_06 does not belong to the four known genotypes, most probably being classifiable into an unrecognized genotype (Fig. 1).

In an attempt to improve the phylogenetic analysis, a maximum-likelihood phylogenetic tree including the 146 full or near-full HEV genomes [reported 145 isolates (Supplementary Table S2) and wbJOY_06] was constructed with the PHYML method version 3.0 (Guindon & Gascuel, 2003) implemented via the web server PALM (http://palm. iis.sinica.edu.tw) (Chen et al., 2009) using the GTR +1 +G substitution model selected by MODELLTEST version 3.7 (Posada & Crandall, 1998) under the Akaike information criterion (AIC). The phylogenetic tree was visualized by FigTree program version 1.2.3 (http://tree.bio.ed.ac.uk/ software/figtree/). The tree further confirmed that wbJOY_06 is distantly related to all known HEV isolates of genotypes 1–4 (Fig. 2).

To investigate possible recombination in the wbJOY_06 genome, a window scanning analysis of aligned HEV genomes was performed using the Simplot software program (version 3.5.1) (Lole et al., 1999). The wbJOY_06 isolate was slightly closer to genotype 4 than the remaining three genotypes, but no significant evidence of recombination between genotype 4 and the other three genotypes was revealed by this method (Supplementary Fig. S1).
Similar to the genomes of genotype 4 (Takahashi et al., 2002; Wang et al., 2000), the wbJOY_06 genome possessed a nucleotide insertion of U between the second and third AUG codons in ORF2, which may change the downstream reading frames in both ORF2 and ORF3 (Fig. 3). ORF2 and ORF3 proteins of genotype 1 HEV have been shown to be encoded by a single bicistronic subgenomic RNA of 2.2 kb in size (Graff et al., 2006), and that initiation of translation at the third in-frame AUG codon of ORF3 in the genotype 1 genome is essential for viral infectivity in vivo (Graff et al., 2005; Huang et al., 2007). The exact initiation site of the subgenomic RNA has also been determined for genotype 3.

**Fig. 1.** An unrooted phylogenetic tree constructed by the neighbour-joining method based on the entire or near-entire nucleotide sequence of 145 HEV isolates of genotypes 1–4 retrievable from the DDBJ/GenBank/EMBL database as of November 2010 (see Supplementary Table S2) as well as the wbJOY_06 isolate obtained in the present study. Since the length of the sequence that has been identified for different isolates varies, the overlapping 7127 nt sequence of 145 reported human, swine, boar, deer, mongoose and rabbit HEV isolates and the wbJOY_06 isolate, which is indicated in boldface type for visual clarity, are compared. Bootstrap values are indicated for the major nodes as a percentage of the data obtained from 1000 resamplings.
HEV (JE03-1760F) and genotype 4 HEV (HE-JF5/15F) in cultured cells, and found to be located exclusively at nt 5122, with a common 5′-GC sequence, which is identical to that of the prototype genotype 1 HEV (Sar-55) (Ichiyama et al., 2009). These results suggest that, despite the presence of an insertion in the genomes of genotype 4, all four genotypes of HEV possess a subgenomic RNA encoding ORF2 and ORF3 proteins with the same N-terminal position. To confirm this notion, the 5′-terminal sequence of the subgenomic RNA of the wbJOY_06 isolate was determined by 5′ RLM-RACE utilizing a procedure specific for capped RNA in accordance with a method described previously (Ichiyama et al., 2009). A sequence analysis of a single product of the expected size (nt 5171–5206: primer sequences at both ends excluded) corresponding to the 2.2 kb subgenomic RNA, revealed that the wbJOY_06 isolate harboured subgenomic RNA with the 5′-terminal GC sequence starting at nt 5171 (corresponding to nt

Fig. 2. Maximum-likelihood phylogenetic relationships of the overlapping 7127 nt sequence of 145 reported human, swine, boar, deer, mongoose and rabbit HEV isolates (see Supplementary Table S2) and the wbJOY_06 isolate. The tree was constructed using PHYML (model GTR+I+G), with optimized tree topology and branch lengths, and the numbers associated with tree branches are indicative of the percentage of 10 full maximum-likelihood bootstrap replicates that support the existence of the branches. The wbJOY_06 isolate is indicated in boldface type for visual clarity.
5122 in the Sar-55 genome) (Fig. 3). Therefore, it is beyond doubt that all mammalian HEV strains, including the wbJOY_06 strain identified in the present study possess a subgenomic RNA starting at the same nucleotide position with the common sequence of 5'-GC. This is further supported by the evidence that a cis-reactive element with the sequence UGAUAUAACAUUG (nt 5153–5164), which is located just upstream of nt 5171, is highly conserved among the all known genotype 1–4 genomes as well as in the wbJOY_06 genome, and has been reported to be critical for synthesis of the 2.2 kb subgenomic RNA (Graff et al., 2005, 2006). Of interest, the wbJOY_06 genome had a A-to-G mutation at nt 5180, with conversion from the methionine codon (AUG) to GUG, and was presumed to encode a shortened ORF3 protein of 112 aa (Fig. 3).

Recently, Takahashi et al. (2010) reported a partial 326 nt ORF1 sequence of an HEV isolate designated JBOAR135-Shiz09 (GenBank accession no. AB573435) that had been recovered from a wild boar captured in 2009 in Shizuoka Prefecture located in the central part of Honshu Island, Japan. The JBOAR135-Shiz09 isolate shared only 71.2–78.2 % nucleotide sequence identities with reported HEV sequences of genotypes 1–4, and was proposed to segregate to genotype 5. Of note, the JBOAR135-Shiz09 isolate was only 79.4 % identical to our wbJOY_06 isolate within the reported 326 nt ORF1 sequence. The phylogenetic tree constructed based on the 326 nt sequence indicated that wbJOY_06 and JBOAR135-Shiz09 were distantly related to each other, similar to the relationship between genotypes 1 and 2, thus suggesting that they may be classifiable into two distinct, unidentified genotypes, although the final decision on genotype classification should be made by the International Committee on Virus Taxonomy based on the comparative analysis of the entire genomic sequences.

In conclusion, we identified an HEV isolate (wbJOY_06) that was recovered from a wild boar in Japan which was found to be highly divergent from the known genotype 1–4 HEV isolates by 22.6–27.7 % over the entire genome. A Simplot analysis revealed no significant recombination between the existing HEV strains of genotypes 1–4. We therefore propose the wbJOY_06 isolate as the first member of a previously unidentified genotype. It remains to be determined if the wbJOY_06-like HEV can cross the species barrier and infect humans or other animal species. Additional studies are warranted to elucidate the host

![Fig. 3. Comparison of the HEV sequences containing potential initiation codons for ORF2 and ORF3. A prototype genotype 1 HEV isolate (Sar-55) and two HEV isolates of genotype 3 [wild-type infectious cDNA clone (pJE03-1760F/wt)] (Yamada et al., 2009b) and genotype 4 (HE-JF5/15F) (Tanaka et al., 2009) which were transfected or inoculated into PLC/PRF/5 cells were compared. The reported initiation site of the subgenomic RNA of Sar-55, pJE03-1760F/wt and HE-JF5/15F (Graff et al., 2006; Ichiyama et al., 2009) and that of wbJOY_06 determined in the present study are illustrated by a vertical bar with an arrow. Potential initiation codons of ORF3 are indicated by shaded boxes, and those of ORF2 are shown by open boxes. The inserted U residue that is found in the genotype 4 isolate (HE-JF5/15F) and the wbJOY_06 isolate is marked with a closed triangle. The termination codon (UGA) of ORF1 is overlined. Dashes indicate nucleotides that are identical to the top sequence and forward slashes denote deletion of nucleotides. Nucleotide positions are in accordance with the wbJOY_06 genome.](image-url)
range and species tropism of the wbJOY_06-like HEV and to search for new HEV strains of unrecognized genotypes in humans and other animals.

Acknowledgements

We thank all of the physicians and hunters who were involved in the collection of serum and/or liver specimens from wild boars. This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Ministry of Health, Labour and Welfare of Japan.

References


