Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Gen Virol. Author manuscript; available in PMC 2017 September 07.

Published in final edited form as:

J Gen Virol. 2016 March ; 97(3): 537–542. doi:10.1099/jgv.0.000393.

Proposed reference sequences for Hepatitis E virus subtypes

Donald B. Smith[1,*], **Peter Simmonds**[1], **Jacques Izopet**[2], **Edmilson F. Oliveira-Filho**[3], **Rainer Ulrich**[4], **Reimar Johne**[5], **Matthias Konig**[6], **members of the ICTV *Hepeviridae* study group**, **Shahid Jameel**[7], **Tim J. Harrison**[8], **Xiang-Jin Meng**[9], **Hiroaki Okamoto**[10], **Wim H.M Van der Poel**[11], and **Michael A. Purdy**[12]

[1]University of Edinburgh, Centre for Immunity, Infection and Evolution, Edinburgh, Scotland, UK [2]INSERM, UMR1043, Toulouse F-31300, France [3]Faculty of Veterinary Medicine, University of Liege, Belgium [4]Federal Research Institute for Animal Health, Reims, Germany [5]Federal Institute for Risk Assessment, Berlin, Germany [6]Justus-Liebig-Univeristy, Giessen, Germany [7]The Wellcome Trust/DBT India Alliance, Hyderabad, India [8]University College London, London, UK [9]College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA [10]Division of Virology, Department of Infection and Immunity, Jichi Medical University School of Medicine, Tochigi-ken, Japan [11]Central Veterinary Institute, Wageningen University and Research Centre, Lelystad, The Netherlands [12]Centers for Disease Control and Prevention, National Center for HIV/Hepatitis/STD/TB Prevention, Division of Viral Hepatitis, Atlanta, Georgia, USA

Abstract

The nomenclature of hepatitis E virus (HEV) subtypes in the literature is inconsistent and makes comparison of different studies problematic. We provide a table of complete genome reference sequences for each subtype. The criteria for subtype assignment vary between different genotypes and methodologies, and so a conservative pragmatic approach has been favoured. Updates to this table will be posted on the ICTV website (link). The use of common reference sequences will facilitate communication between researchers and help clarify the epidemiology of this important human pathogen. This subtyping procedure might be adopted for other *Orthohepevirus* taxa.

The current literature contains several inconsistencies in the naming of hepatitis E virus (HEV) subtypes, which often creates confusion in the HEV scientific community. The current taxonomic position of HEV is that it is a member of the family *Hepeviridae* within the genus *Orthohepevirus*. Four species have been defined that infect birds (*Orthohepevirus B)*, rodents, soricomorphs and carnivores (*Orthohepevirus C)*, or bats (*Orthohepevirus D)*. The largest species, *Orthohepevirus A*, comprises seven genotypes that infect human (HEV 1, 2, 3, 4 & 7), pig (HEV- 3 & 4), rabbit (HEV-3), wild boar (HEV-3, 4, 5 & 6), mongoose (HEV-3), deer (HEV-3), yak (HEV-4) and camel (HEV-7) (Smith *et al.*, 2014).

This division of HEV into 7 genotypes and criteria for their assignment and identification are based on a demarcation p-distance threshold between genotypes of 0.088 for amino acid

[*]Corresponding author: D.B.Smith: D.B.Smith@ed.ac.uk Tel: (+44) 131 650 7331.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

distances of concatenated ORF1 and ORF2 (lacking hypervariable regions between ORF1 amino acids 706–778 and 928–929, numbered with reference to M73218) (Smith *et al.*, 2014). However, the criteria by which HEV variants can be assigned to subtypes within genotypes are less consistent and sometimes confusing. When HEV subtypes were first comprehensively tabulated a decade ago only 49 complete genome sequences were available and many subtype assignments were based on the analysis of subgenomic regions (Lu *et al.*, 2006). Since then, the number of complete genome sequences has increased to almost 300 and most of the subtypes defined by Lu et al., 2006 are now represented by at least one complete genome sequence. However, there is currently no agreed list of reference sequences for these subtypes, although an attempt at standardisation has been made for HEV-3 (Smith *et al.*, 2015). One problem that is encountered in assigning sequences to particular subtypes is that no consistent criteria have been identified that define intra- and inter-subtype distances (Oliveira-Filho *et al.*, 2013; Smith *et al.*, 2013). For example, nucleotide p-distances between subtypes of HEV-1 are all less than 0.12, while those between subtypes of HEV-3 range from 0.12 to 0.26 and from 0.13 to 0.18 for subtypes of HEV-4. In addition, within these genotypes, the ranges of within and between subtype distances overlap. As a result, some complete genome sequences have been given conflicting subtype assignments.

An example comes from a recent paper (Lhomme *et al.*, 2015) in which strain TR19 (JQ013794), was used as the reference sequence for subtype 3c. The frequency of subtype 3c infections has increased over the last decade in France, similar to the increase in subtype 3c previously documented in England and Wales (Ijaz *et al.*, 2014). However, the “subtype 3c” strains from the UK actually correspond to the subtype 3i reference sequence used in the French study. In other cases, subgenomic sequences used as reference sequences (Thiry *et al.*, 2015) derive from strains for which no further sequence information is available. As a result, it has become difficult to compare phylogenetic analyses carried out using different subgenomic regions or even the same region in different studies.

To address these issues we propose a standard reference set of complete genome sequences (Table 1). This Table is available online on the ICTV website (link) and will be updated as new information becomes available. The criteria used are as follows:

1. To minimise disruption of the literature, priority was given to the subtype assignments given by Lu et al., 2006.
2. To enable phylogenetic analyses to be carried out on different fragments of the genome, subtype reference sequences must comprise both the ORF1 and ORF2 coding regions and not be a recombinant between previously assigned subtypes.
3. If more than one complete genome sequence was available for a subtype, priority was given to the first sequence to be submitted to GenBank or, where submission dates were identical, the lowest alphabetic/numeric Accession number.
4. If a subtype was assigned by Lu et al., 2006 based on the analysis of subgenomic fragments, these fragments were used to identify potential reference sequences by performing a BLAST search against GenBank. The highest scoring complete genome sequences were considered as potential reference sequences if BLAST

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

scores were >90% and if sequence identities formed a discontinuous distribution compared to scores for previously named complete genome sequences.

5. Complete genome sequences that were phylogenetically distinct from previously assigned complete genome sequences and not related to any of the subtypes described by Lu et al., 2006 were only assigned as a new subtype if at least three complete genome sequences were available that were epidemiologically unrelated (from different studies or localities). Unassigned complete genome sequences were labelled “genotype_Accession number” (e.g. “3_ AB369689”)

We considered an alternative method in which the most central sequence (the medoid) in each subtype group would become the reference sequence. Although not without advantages, this method would also mean that subtype reference sequences would not be stable because the medoid may change as more sequences are obtained or as the structure of the subtype is redefined by the addition or exclusion of divergent strains. In addition, our decision to use the designations of Lu et al., 2006 with priority to strains with the earliest date of Accession will be minimally disruptive to the existing literature.

Phylogenetic and sequence analyses

HEV sequences > 7000 nucleotides long were downloaded from GenBank database on 27th October 2015 and aligned using SSE v1.2 (Simmonds, 2012). Sequences differing by <1% (HEV-1 and HEV-3) or 2% (HEV-4) of nucleotide positions were analysed by producing neighbour joining trees, based on maximum composite likelihood distances, using MEGA6 (Tamura *et al.*, 2013), or by analysing the distribution of nucleotide p-distances using SSE. Analyses in sequence sets lacking hypervariable regions or lacking the overlapping ORF2/3 region produced similar results.

Genotype 1

Subtypes 1a–1e were all originally assigned on the basis of an analysis of complete genome sequences (Lu *et al.*, 2006). A group of sequences that share a common branch with subtype 1a (JF443721-26 and AB720035) are more divergent from subtype 1a (nucleotide distances 0.052–0.075, apart from M73218 to JF443726, 0.046) than sequences of subtype 1a are from each other (<0.056), these distances being comparable to those between subtypes 1b and 1c (0.058–0.065). We propose that this phylogenetically distinct group of sequences be considered as subtype 1f, although no discontinuity exists in the distribution of pairwise nucleotide p-distances within HEV-1 sequences that distinguishes within and between subtype distances, Sequence FJ457024 is intermediate between subtypes 1a and 1f, but bootscan analysis using SSE suggests that it is a recombinant between these two subtypes (data not shown). All p-distances greater than 0.087 derive from comparisons between subtypes 1a, 1b, 1c and 1f and subtypes 1d and 1e (>0.101), or between subtypes 1d and 1e (0.096), supporting the division of HEV-1 into two clades: 1abcf (comprising subtypes 1a, 1b, 1c and 1f) and 1de (subtypes 1d and 1e).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Genotype 2

Only a single complete genome sequence has been reported for genotype 2a; genotype 2b was identified from the analysis of a 318 nt ORF2 fragment.

Genotype 3

The distribution of nucleotide distances amongst HEV-3 subtypes shows a complex pattern with multiple hierarchies of relatedness, even if the more divergent rabbit-derived strains are excluded. Subtypes 3a, 3b, 3c, 3h, 3i and 3j (3abchij) form one major clade, while subtypes 3e, 3f and 3g form another (3efg) (Hewitt *et al.*, 2014; Ijaz *et al.*, 2014; Oliveira-Filho *et al.*, 2013; Smith *et al.*, 2015; Widén *et al.*, 2011). The reference sequences for subtypes 3e and 3f were assigned according to date of Accession to Genbank. Five strains belonging to subytpe 3c were listed by Lu et al., 2006; their partial ORF1 and ORF2 sequences group with the corresponding regions of the complete genome sequence FJ705359, and separately from JQ013794, previously described as subtype 3c (Izopet *et al.*, 2012). The latter sequence becomes the subytpe 3h reference sequence since it groups with the ORF1 and ORF2 sequences of a subtype 3h strain listed by Lu et al., 2006., (AF110390, AF110387). The other 3h strain (swNZ) listed by Lu et al., 2006 groups separately from all complete genome sequences. Four strains of subtype 3i are listed by Lu et al., 2006; sequences of one of these strains groups with FJ998008 for both the ORF1 and ORF2 regions. The other three strains have sequences only loosely (ORF1) or not associated (ORF2) with this sequence. Accordingly we have assigned FJ998008 as the 3i reference sequence. Nucleotide p-distances between these subtypes (>0.120) overlap distances within subtypes (<0.123) making it difficult to unambiguously assign some subtypes. For example, nucleotide p-distances between subtype 3f and EU360977 (0.116 to 0.125) and between 3f and KJ873911 (0.116 to 0.125) span this range as does that between 3h and AB290312 (0.120), while AB369689 and AB740232 are equally related to subtypes 3a (nucleotide p-distances 0.124–0.134) and 3b (0.126–0.137). We have chosen not to assign a subtype to these sequences, or to more divergent sequences such as JQ953664, AB290313. Divergence amongst the HEV-3 rabbit-derived strains range up to 0.255, again with multiple levels of sequence divergence; assignment of these strains into subtypes within the 3ra clade awaits the availability of further complete genome sequences.

Genotype 4

Seven HEV-4 subtypes were defined by Lu et al., 2006 (subtypes 4a to 4g). The distribution of nucleotide sequence distances between and within HEV-4 subtypes is nearly continuous with distances between subtypes (> 0.133) overlapping those within subtypes (<0.139) although a peak from 0.15 to 0.18 consists only of distances between subtypes. Phylogenetic analysis also reveals multiple levels of branching (Figure 1) but without higher level groupings akin to those observed for HEV-1 and HEV-3. Consequently, we have used a pragmatic approach, adopting previous designations and avoiding the proliferation of new subtype names. One of the 4f subgenomic ORF1 accession numbers given by Lu et al., 2006 (AY427953) should be AY684253. However, both this sequence and another subtype 4f ORF1 sequence (AB075970) group with the subtype 4a reference sequence. Two additional

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

subtype 4f sequences given by Lu et al., 2006 (AB082547 and AB082558) derive from the HE-JA2 strain for which a complete genome sequence is now available (AB220974) and which is distinct from previously named subtypes, so this becomes the subtype 4f reference sequence. Two additional subtypes (4h and 4i) follow the assignments given in a previous publication (Liu *et al.*, 2012). Sequence AB369688, although distinct from other subtypes, is represented by a single complete genome sequence and therefore remains unassigned.

Genotypes 5–7

The distance between the two complete genome sequences of HEV-6 (AB602441 and AB856243) is 0.198, and between the three complete genome sequences of HEV-7 (KJ496143, KJ496144 and KT818608) is 0.06–0.147. Comparison with distances between subtypes of HEV-3 and between subtypes of HEV-4 would suggest that both HEV-6 and HEV-7 could also be divided into two subtypes. However, as fewer than three complete genome sequences are currently available of each variant we have not made any subtype assignments except to designate the first sequence of each genotype as subtype “a”.

Concluding remarks

A perennial problem in classifying virus diversity is that discrete, man-made categories used for classification become arbitrary as their genetic distinctness blurs into a continuum of variability with the description of additional novel variants or recombinants. This problem has hindered the assignment of subtypes of HEV because of different levels of diversity within different HEV genotypes and because both distance based and phylogenetic methods do not provide clear criteria for demarcation between groups. Despite this problem, it is important that researchers have a common set of named reference sequences, so that results from different studies can be compared. We hope that our table of subtype reference sequences will assist the interpretation of epidemiological and evolutionary studies of HEV. However, an important caveat is that researchers should use these reference sequences as way-markers in a complex landscape and be cautious about treating subtypes as stable biological or epidemiological entities.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Hewitt PE, Ijaz S, Brailsford SR, Brett R, Dicks S, Haywood B, Kennedy ITR, Kitchen A, Patel P, et al. Hepatitis E virus in blood components: a prevalence and transmission study in southeast England. Lancet. 2014; 384:1766–73. [PubMed: 25078306]

Ijaz S, Said B, Boxall E, Smit E, Morgan D, Tedder RS. Indigenous Hepatitis E in England and Wales From 2003 to 2012: Evidence of an Emerging Novel Phylotype of Viruses. J Infect Dis. 2014; 209:1212–1218. [PubMed: 24273173]

Izopet J, Dubois M, Bertagnoli S, Lhomme S, Marchandeau S, Boucher S, Kamar N, Abravanel F, Guérin JL. Hepatitis E virus strains in rabbits and evidence of a closely related strain in humans, France. Emerg Infect Dis. 2012; 18:1274–81. [PubMed: 22840216]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lhomme S, Abravanel F, Dubois M, Chapuy-Regaud S, Sandres-Saune K, Mansuy JM, Rostaing L, Kamar N, Izopet J. Temporal evolution of the distribution of hepatitis E virus genotypes in Southwestern France. Infect Genet Evol. 2015; 35:50–5. [PubMed: 26218544]

Liu P, Li L, Wang L, Bu Q, Fu H, Han J, Zhu Y, Lu F, Zhuang H. Phylogenetic analysis of 626 hepatitis E virus (HEV) isolates from humans and animals in China (1986–2011) showing genotype diversity and zoonotic transmission. Infect Genet Evol. 2012; 12:428–434. [PubMed: 22306814]

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. [PubMed: 16175650]

Oliveira-Filho EF, König M, Thiel HJ. Genetic variability of HEV isolates: inconsistencies of current classification. Vet Microbiol. 2013; 165:148–54. [PubMed: 23434129]

Simmonds P. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res Notes. 2012; 5:50. [PubMed: 22264264]

Smith DB, Purdy MA, Simmonds P. Genetic variability and the classification of hepatitis E virus. J Virol. 2013; 87:4161–9. [PubMed: 23388713]

Smith DB, Simmonds P, Jameel S, Emerson SU, Harrison TJ, Meng XJ, Okamoto H, Van der Poel WHM, Purdy MA. Consensus proposals for classification of the family Hepeviridae. J Gen Virol. 2014; 95:2223–32. [PubMed: 24989172]

Smith DB, Ijaz S, Tedder R, Hogema B, Zaaijer H, Izopet J, Bradley-Stewart A, Gunson R, Harvala H, et al. Variability and pathogenicity of Hepatitis E virus genotype 3 variants. J Gen Virol. 2015; 96:3255–64. [PubMed: 26282123]

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30:2725–9. [PubMed: 24132122]

Thiry D, Mauroy A, Saegerman C, Licoppe A, Fett T, Thomas I, Brochier B, Thiry E, Linden A. Belgian Wildlife as Potential Zoonotic Reservoir of Hepatitis E Virus. Transbound Emerg Dis. 2015

Widén F, Sundqvist L, Matyi-Toth A, Metreveli G, Belák S, Hallgren G, Norder H. Molecular epidemiology of hepatitis E virus in humans, pigs and wild boars in Sweden. Epidemiol Infect. 2011; 139:361–71. [PubMed: 20546635]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

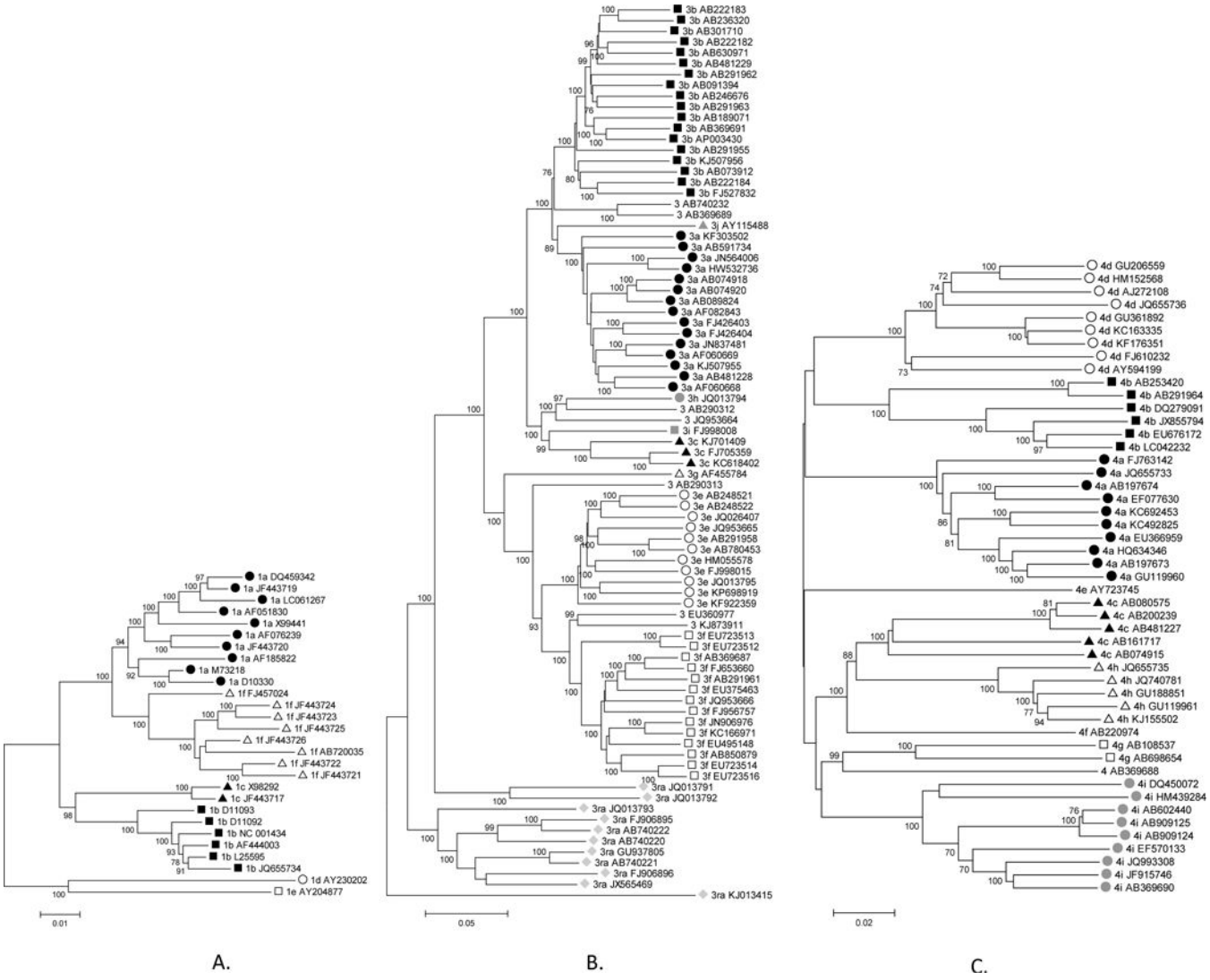

Figure 1.

Phylogenetic analyses of HEV complete genome sequences. A neighbour joining tree of maximum likelihood distances is shown with symbols used to indicate sequences belonging to the same subtype of A. HEV-1, B. HEV-3, and C. HEV-4. Branches supported by > 70% of bootstrap replicates are indicated.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Genotype	Subtype	Accession	Strain	Subgenomic reference sequences/comments
1	1a	M73218	Burma	
	1b	D11092	HPECG	
	1c	X98292	I1	
	1d	AY230202	Morocco	
	1e	AY204877	T3	
	1f[1]	**JF443721**	**IND-HEV-AVH5-2010**	
2	2a	M74506	M1	
	2b[2]			AF173231-2, AY903950 (ORF2)
3	3a	AF082843	Meng	
	3b	AP003430	JRA1	
	3c	**FJ705359**	**wbGER27**	
	3d[2]			AF296165-7 (ORF2)
	3e	**AB248521**	**swJ8-5**	
	3f	**AB369687**	**E116-YKH98C**	
	3g	AF455784	Osh 205	
	3h	**JQ013794**	**TR19**	
	3i	**FJ998008**	**BB02**	
	3j	AY115488	Arkell	Isolated from pooled material
	3[3]	AB290312	swMN06-A1288	
	3	JQ953664	FR-SHEV3c-like	
	3	AB369689	E088-STM04C	
	3	AB290313	swMN06-C1056	
	3	EU360977	swX07-E1	
	3	KJ873911	FR_R	
	3	EU723513	SW627	
	3ra	**FJ906895**	**GDC9**	Mostly from rabbit, includes several subtypes
4	**4a**	**AB197673**	**JKO-ChiSai98C**	
	4b	**DQ279091**	**swDQ**	
	4c	AB074915	JAK-Sai	
	4d	AJ272108	T1	
	4e	**AY723745**	**IND-SW-00-01**	
	4f	**AB220974**	**HE-JA2**	
	4g	AB108537	CCC220	
	4h	**GU119961**	**CHN-XJ-SW13**	
	4i	**DQ450072**	**swCH31**	
	4	AB369688	E087-SAP04C	
5	**5a**	**AB573435**	**JBOAR135-Shiz09**	From wild boar

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Genotype	Subtype	Accession	Strain	Subgenomic reference sequences/comments
6	**6a**	**AB602441**	**wbJOY_06**	From wild boar
	6	AB856243	wbJNN_13	From wild boar
7	**7a**	**KJ496143**	**178C**	From camel
	7	KJ496144	180C	From camel

[1]Reference sequences not assigned a subtype by Lu et al., 2006 are highlighted by bold text.

[2]Subtypes 2b and 3d are defined from Lu et al., 2006 by the subgenomic sequences indicated.

[3]Unassigned subtypes are denoted by genotype without a subtype designation.