Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Virological Methods

Full coding hepatitis E virus genotype 3 genome amplification method

M. Muñoz-Chimeno [a], J.E. Forero [b], J.M. Echevarría [a], J.L. Muñoz-Bellido [c],
L. Vázquez-López [d], L. Morago [a], M.C. García-Galera [a], A. Avellón [a,*]

[a] Hepatitis Unit, National Center of Microbiology, Instituto de Salud Carlos III, Madrid Spain
[b] Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Medellín Colombia
[c] Microbiology Unit, Salamanca University Hospital, Salamanca Spain
[d] Hematology Unit, Salamanca University Hospital, Salamanca Spain

A B S T R A C T

Hepatitis E virus (HEV) genotype 3 produces zoonotic infection associated with the consumption of infected animals. HEV infections can become chronic in immunocompromised (IC) patients. The viral genome has three well defined open reading frames (ORF1, ORF2 and ORF3) within which various domains and functions have been described. This paper (i) describes a new method of complete sequencing of the HEV coding region through overlapping PCR systems, (ii) establishes a consensus sequence and polymorphic positions (PP) for each domain, and (iii) analyzes the complete coding sequence of an IC patient. With regard to the consensus, a high percentage of PP was observed in protease (PP = 19%) and the X domain (PP = 22%) within ORF1, the N-terminal region of the S domain (PP = 22%) in ORF2, and the P1 (PP = 35%) and P2 (PP = 25%) domains in ORF3. In contrast, the ORF1 Y, ORF2 S, ORF2 M and ORF3 D1 domains were conserved in the reference sequences (0.40, 1, 0.70 and 0% of PP, respectively). The sequence from the IC patient had more mutations in the RpRp (D1235G, Q1242R, S1454T, V1480I, I1502 V, K1511R, G1373 V, E1442D, V1693 M), the terminal ORF2 S- domain (F10L, S26T, G36S, S70P, A105 V, I113 V), the X domain (T938 M, T856 V, S898A) and the helicase (S1014N, S975T, Q1133 K).

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Hepatitis E virus (HEV) (Hepeviridae family) is one of a number of divergent isolates from humans and other animals (Song et al., 2014) whose taxonomic status is unresolved. This infection is endemic in many parts of the US, Central and parts of South America, Mediterranean regions of Europe, Africa and Asia Pacific region (Perez-Gracia et al., 2013). The clinical course of genotype 3 (G3) infection is usually self-limited in immunocompetent individuals (Rodriguez-Frias et al., 2012). In contrast, chronic and fulminant infections in patients coinfected with human immunodeficiency virus (Dalton et al., 2011), with iatrogenic immunosuppression or with solid organ transplantation (Kamar et al., 2008) have been described as well as severe acute infections in pregnant women (Mateos Lindemann et al., 2010). In fact, 60% of HEV infections

in immunocompromised patients after solid organ transplantation evolve to chronic infection without antiviral treatment (Fujiwara et al., 2014). HEV infection has been reported to be associated with allogeneic hematopoietic transplantation (Versluis et al., 2013), or with transplantation of solid organs such as the liver (Te et al., 2013), kidney (Passos et al., 2013; Moal et al., 2013) and heart (Koning et al., 2013). Cases of HEV transmission due to transfusions (Price, 2014) have also been noted. A case has recently been described from Germany in which a low HEV G3 viral load was transmitted by platelets to an immunocompromised patient who subsequently developed chronic hepatitis (Huzly et al., 2014).

The HEV genome is an ss+RNA of approximately 7.2 kb, including three partially overlapping open reading frames, from 5′ to 3′: ORF1, ORF2 and ORF3, these structures are thought to be important for HEV RNA replication (Reyes et al., 1993). ORF1 encodes a 1693 amino acids (AAs) long polyprotein that has structural and functional motifs responsible for viral replication: methyltransferase, papain-like cysteine protease (PCP), helicase and RNA-dependent RNA polymerase (RdRp) (Ahmad et al., 2011). However, some uncharacterized domains homologous with other animal and plant RNA positive-strand viruses are also found within the ORF1 protein. These include the Y domain, the X domain (formerly the macro

* Corresponding author.
    E-mail addresses: milagrosmch@hotmail.com (M. Muñoz-Chimeno),
jforeroduarte@gmail.com (J.E. Forero), jmecheva@isciii.es (J.M. Echevarría),
jlmubel@usal.es (J.L. Muñoz-Bellido), lvazlo@usal.es (L. Vázquez-López),
lmorago@isciii.es (L. Morago), Mcgarciagalera@gmail.com (M.C. García-Galera),
aavellon@isciii.es (A. Avellón).

domain) and a polyproline or hypervariable region (Holla et al., 2013). ORF2 encodes a 660 AAs protein which is the main component of the viral capsid (Zafrullah et al., 1999). The capsid is formed by the interaction of ORF2 protein dimers with the 5′ end of the viral genome through its N-terminal region, which is rich in arginine (AA 101; 180 molecules of proORF2 per virus particle) (Xing et al., 2010). This protein has three linear domains: the S domain, which forms the capsid, and the M and P domains, which are involved in the interaction of the virus with the host cell (Perez-Gracia et al., 2014). It has been described an homologous to alphavirus region (46 AAs long) of unknown function (Yarbough et al., 1991; Purdy et al., 1993; Rodriguez-Frias et al., 2012). The ORF3 protein probably regulates the host cell environment through its interaction with various intracellular pathways. Firstly, it activates the ERK by binding and inhibiting their phosphatase. Prolonged activation of ERK would generate a survival and proliferative cellular sign that can favor viral replication and extend the life of infected cells by attenuating the intrinsic pathway of death. Secondly, the innate immune response is downregulated by reducing acute phase protein expression and increasing secretion of α1-microglobulin with immunosuppressive capacity (Chandra et al., 2008). The ORF3 protein interacts with ORF2. This interaction is dependent upon the phosphorylation of serine 71, and suggests a role for ORF3 protein in regulating the assembly of HEV and the possible location of the virus capsid (Tyagi et al., 2002).

Reports of molecular analysis of HEV in clinical case samples are scarce and they are focused in specific genome regions (Lhomme et al., 2014b). The aim of this work is to analyze the complete coding genome of the HEV genotype 3 in an immunocompromised patient with acute infection. For this purpose, a new full coding genome amplification method is described, and a comprehensive analysis of polymorphic residues of a Genbank HEV complete sequence set is carried out.

2. Materials and methods

2.1. Samples

RNA extracts from serum and feces of a patient with acute HEV genotype 3 hepatitis were used as a template for cDNA synthesis (hereafter, the "IC sample"). The patient had not recently traveled outside Spain and was immunocompromised after bone marrow transplantation. Clinical presentation was acute hepatitis with neither persistence nor evidence of any serological (anti HEV IgM or IgG) response. After other virological causes of hepatitis were ruled out, HEV PCR was determined with a positive result.

2.2. Laboratory methods

2.2.1. RNA extraction

RNA was extracted manually. RNA was extracted from the stools with QIAamp® Viral RNA Mini Kit (QIAGEN®, Hilden, Germany), which simplifies purification of viral RNA with a fast spin-column. Stool samples were homogenized and resuspended in PBS 1X (1.125 ml). RNA was extracted from serum samples (200 μl) with a High Pure Viral Nucleic Acid kit (Roche Diagnostics, Mannheim, Germany). The method uses chaotropic salt and glass fiber fleece in a high pure spin filter tube.

2.2.2. Obtaining cDNA

This was done with a Transcription First Strand cDNA Synthesis Kit (Roche Diagnostics, Mannheim, Germany) adding 10 μl of RNA extract to a final amount of 20 μl cDNA, following the manufacturer's recommendations.

2.2.3. Amplification strategy and primer design

To amplify the HEV genome coding region, 12 overlapping blocks of nested PCR systems were set up. An additional block was used with pre-designed nucleotides (Avellón et al., 2015). 18 and 24 oligonucleotides were designed for the initial and nested reactions, respectively. Primer sequences and thermal conditions are summarized in Table 1. To design the primers, 70 complete HEV genotype 3 genome sequences were obtained (GenBank database, National Center of Biotechnology Information, Bethesda, MD, USA) and aligned (Mega 5.0 software). Primers were designed with the help of Seqman software (DNASTAR, Lasergene Inc., Madison, WI, USA).

2.2.4. Polymerase chain reaction amplification and product detection

PCR Master Mix and RNase-free water were used (Promega, Madison WI, USA) to perform first (from 5 μl of extracts) and nested (from 2.5 μl of first PCR amplification product) PCR reactions. Previously designed sense and antisense primers were used at a concentration of 0.5 μM. Amplification was carried out in a Thermocycler (DNA Engine, MJ Research PTC-200) programmed as follows: 2 min at 94 °C (initial denaturation); 40 cycles: 1 min at 94 °C (denaturation), 1 min annealing (temperature of each system described in Table 1), 1 min at 72 °C (elongation): 5 min at 72 °C (final elongation). PCR products were separated by electrophoresis on 2% agarose gels in Tris-borate-EDTA (TBE 10X) buffer, and stained (Biotium GelRed®, Hayward, CA, USA).

2.2.5. Amplification product purification

The amplification products were purified with a QIAquick PCR Purification Kit (QIAGEN®, Hilden, Germany)

2.3. Sequence analysis

2.3.1. Sequencing and sequence assembly

Sense and antisense DNA strands were both sequenced by the Sanger method. Sequence assembly and AA alignment were performed with Seqman and MegaAlign programs (DNAS-TAR, Lasergene Inc.). The sequence obtained (hereafter, the "IC sequence") was submitted to the NCBI database under accession number KU513561.

2.3.2. Genbank HEV complete sequence set

62 complete HEV genotype sequences from the GenBank database were used as references (hereafter mentioned as "reference sequences"): AB369691, AB369689, AB291951, AB369687, AB291953, AB291957, FJ653660, AB189071, AB437317, AB593690, JN837481, AB290312, FJ956757, AB301710, AB291955, AB291954, EU723516, AB189074, AB443626, AB189072, EU723513, AB189073, EU723515, AB189075, AB291952, AB291956, AB443627, AB443623, AB443624, FJ998008, AB362841, AB630971, AB189070, JQ679013, AB291961, HQ389544, AB362840, HQ709170, AB437316, FJ527832, HQ389543, AB291960, EU360977, JQ953664, AB740232, AB362842, JQ679014, AB362839, AB222184, AB437318, AB291962, AB246676, AB425830, AB248522, EU495148, AB630970, AB425831, AB248521, FJ705359, AB291963, AB222182 and AB222183.

2.3.3. Heterogeneity study of reference sequences

The degree of conservation of viral proteins was determined by calculating the percentage of polymorphic positions (PPs) in each protein. A residue was considered to be in a PP when mutations in more than one reference sequence were found. The PP% was calculated as PP $\times$ 100/number of total amino acids in each protein. The

Table 1
Oligonucleotides used for amplification of complete HEV coding genome initial reaction (1st R) and nested reaction (2nd R). Tm: annealing temperature. I: inosine.

1st R	B1	B2	B3	B4	B5	B6	B7	B8	B9
Name	E3_12S	E3_561S	E3_22S	E3_1527S	E3_2032S	E3_3475S	E3_4759S	E3_5236S	E3_6474S
Sequence (5′-3′)	ACGYATGTGGT CGAWGCCATG	ATGGCCCGICA YGGGATGAC	TCGAWGCCAT GGAGGCCCA	GGCYTAYGAR GGITCIGARGT	TRTGGYTRCA YCCYGAGGG	CTYGGYGAYC CIAAYCAGAT	ATGGARGART GYGGYATGCC	TGCCTATGYTG CCCGCGC	GCIGCYACRCGI TTYATGAA
Name	E3_987A	E3_2091A	E3_1584A	E3_3025A	E3_3781A	E3_5236A	E3_5597A	E3_6648A	E3_7371A
Sequence (5′-3′)	AARAGCATRA GCCGRTCCCA	AAIGGRTTIG CIGAYTCCCA	CCRTGRACIGC RTARGTCCC	GGGCGRTGRT TRCGYTCCCA	CTRGCRTCRG CYGTRGCTAT	ACATCRACAC ARACCTGCGC	GGGTTGGTTGGA TGAATATAGGGG	ACWGYYGGCT CACCATTGGC	TTYTAAGRCGC TGAAGYTCAG
Size (nt)	975	1530	1562	1498	1749	1761	838	1412	897
Tm (°C)	60	60	60	57	57	50	57	57	57

2nd R	B1n	B2n	B3n	B4n_1	B4n_2	B5n_1	B5n_2	B6n_1	B6n_2	B7n	B8n	B9n
Name	E3_22S	E3_984S	E3_120S	E3_1527S	E3_2032S	E3_2896S	E3_3316S	E3_3592S	E3_4192S	E3_4895S	E3_5478S	E3_6587S
Sequence (5′-3′)	TCGAWGCCA TGGAGGCCCA	ATYTGGGAYC GGCTYATGCT	GCTGTGGTGG TYCGGCCGTT	GGCYTAYGAR GGITCIGARGT	TRTGGYTRCA YCCYGAGGG	AGGRTYGARC AGAAYCCIAA-GAG	GAGCTYCGIAA YAGYTGGCG	GAYGTITGYGA RCTYATACG	ATAGTYCAYTG YCGIATGGC	GGAAYACYGTY TGGAAYATGGC	TGCCTATGYTG CCCGCGC	CCGACAGAATT RATTTCGTCGGC
Name	E3_561A	E3_1584A	E3_1314A	E3_2091A	E3_2923A	E3_3563A	E3_3781A	E3_4384A	E3_5227A	E3_5592A	E3_6631A	E3_7302A
Sequence (5′-3′)	GTCATCCCRT GICGRGCCAT	CCRTGRACIG CRTARGTCCC	CGRCAYTGIG CRTARAACTG	AAIGGRTTIG CIGAYTCCCA	GTYTCICGRT AYGCYGCCTC	CGRTGIGTIA CRTGCCACCA	CTRGCRTCRG CYGTRGCTAT	TCCTGRCCCT TYTCIACCAT	CARACCTGCG CIACRTTCGT	TTGGTTGGATG AATATAGGGGA GGG	CCCGCGC GGCYGARACRA CIGGGCGGG	TCMGGRCARAA ATCATCRAAAGT
Size (nt)	539	600	1194	564	891	667	465	792	1035	697	1076	715
Tm (°C)	60	60	57		54		54		50	57	57	57

degree of conservation was arbitrarily established as: high, PP% ≤ 1 *vs.* low PP% ≥ 15.

2.3.4. Consensus definition

This was obtained from reference sequences as follows: the amino acid consensus sequence was established for each protein according to the most frequent amino acid in each position. The consensus sequence was used to determine the mutation pattern of the IC sample, excluding the polymorphic positions.

3. Results

The PP% values of each protein are summarized in Tables 2 (ORF 1) and 3 (ORF 2 and ORF 3). With regard to ORF1, the Y domain was highly conserved (PP% = 0.4), while the protease and X domain had a low level of conservation (PP% = 19 and 22, respectively). Protease catalytic site residues were Cys-Tyr and Cys-Cys in 53 and 9 sequences, respectively. RdRp-conserved motifs III and IV presented several PPs. In ORF2, the C-terminal region of the S and M domains was highly conserved (PP% = 1 and 0.70, respectively) while the N-terminal region of the S domain had a low level of conservation (PP% = 22). Finally, analysis of ORF3 revealed an extremely high degree of conservation in the D1 domain (PP% = 0) while the other domains (D2, P1 and P2 domains) presented low conservation levels (PP% = 16, 35 and 25, respectively). The polyproline region was the most variable region in the genome. A large number of reference sequences (*n* = 43) were 110 AAs long, but we found sequences with insertions (*n* = 8) and deletions (*n* = 11) in this region. The homologous alphavirus region harbored T16C and T19G mutations in 53% and 89% of reference sequences, respectively.

With regard to the IC sample, sequences obtained from serum and feces were identical, thus results presented in tables are applicable to sequences obtained of both samples. We found more than one mutation in the X domain, helicase and RdRp of ORF1 polyprotein (Table 2). The PCP catalytic site was Cys-Tyr, and the RdRp conserved domains I, II, V–VIII were fully conserved as well as the N-terminal region of the S domain of the ORF2 protein (Table 3). The IC sequence showed neither deletions nor insertions in the polyproline region. The homologous alphavirus region had T16C and T19G mutations.

4. Discussion

In the nearly 40 years since the discovery of HEV, our knowledge of its genome and life cycle has grown only slowly. Despite its undeniable epidemiological and clinical significance, some technical issues, such as its adaptation to culture *in vitro*, have presented obstacles to researchers. Since Koonin and colleagues (Koonin et al., 1992) first described its genomic structure in 1992, there have been no significant advances in our understanding of the genome structure; the sequence published by this group is still used as a reference by many researchers. Thus, our method should contribute to the knowledge of the HEV genome, as it enables direct and sensitive amplification of the complete coding genome from clinical samples, unlike most published methods, which have been standardized for use after HEV has been grown in cell culture e.g. (Tanaka et al., 2007). Our system would allow clinical samples (of serum or feces) to be used to analyze the complete sequence in human HEV infections of particular interest to public health; for instance comparative genome analysis of chronic and acute infections could be done. Furthermore, it is important to have highly sensitive research tools to investigate HEV transmission by blood products, since this is known to be possible with low viral loads (Huzly et al., 2014). To date, the only complete HEV genome sequences available in databases from Spanish research groups

correspond to viruses found in pigs (EU723513, EU723516 and EU723515) (Peralta et al., 2009). There are few reports in which the HEV genome has been analyzed in immunocompromised patients (Lhomme et al., 2014b).

A crucial role in the viral cycle was proposed for methyltransferase, PCP and helicase proteins (Holla et al., 2013). From the analysis of our reference sequences, we have brought to light that methyltransferase, PCP and helicase present numerous polymorphic AA positions. Viral methyltransferase is essential for HEV replication and infectivity, and is required for efficient binding of mRNA to ribosomes as well as for avoiding the innate immune response mediated by interferon (Pichlmair et al., 2006). Heterogeneity of the second AA of the PCP putative catalytic site (Tyr/Cys), imprecision in the function and limits for the region that has been proposed but not confirmed as being a functional protease (Suppiah et al., 2011), together with its low level of conservation, highlights the need for functional studies in relation to this protein. Conversely, conservation of RdRp I–VIII motifs including the GDD sequence that binds Mg²⁺, which is essential for replication, is consistent with its proposed key functional role (Koonin, 1991). Two described domains are still not functionally characterized at present. Of these, the Y domain is highly conserved in our analysis, suggesting that an important function is located in this region, although little is known about it. In the case of the X or macro domain, although its function is unclear, it is analogous to other viral genomes that are better known as coronaviruses (Koonin et al., 1992). It may not be essential for replication, given its high tolerance to polymorphic AAs. The macro domain is known to be involved in the inflammation process in animal models, and influences viral pathogenicity by substitution of a conserved AA residue that is responsible for the reduced secretion of inflammatory cytokines. However, this function has not been described in the case of HEV (Lhomme et al., 2014b).

The proline-rich region was confirmed as being the most variable genomic region of the virus. A recent study showed the mutation rate of the proline-rich region to be similar to that of other ORF1s, but due to a higher substitution rate and a preference for cytosine in the first and second positions, there is a higher frequency of proline residues in this region (Holla et al., 2013). This region seems to be of special interest because of its potential relationship with the infectivity of HEV G3 in various animal species (Purdy et al., 2012), as well as because of its possible relationship with immune control and viral persistence (Lhomme et al., 2014b). Future studies of cases of chronic infection could provide useful information about the significance of this region. The region that is homologous to the alphavirus and that contains the ORF1 stop codon and the ORF2 and ORF3 initial codons is highly conserved (Rodriguez-Frias et al., 2012). Changes were observed in two nucleotides, none of which was related to the aforementioned codons.

According to our results the ORF2S and M domains are conserved in its C-terminal extreme. The S domain is considered to be a globally conserved region among the different HEV genotypes but has greater divergence in the N-terminal region. The M domain strongly interacts with the S domain and connects the P domain through a proline-rich region (Xing et al., 2010). In contrast, ORF2P domain presents greater variation, in possible association with its putative function and the presumed immune escape as is has been related to cell entry and antibody neutralization as well as a target for some of the vaccines currently under development (Ahmad et al., 2011; Li et al., 2009).

ORF3 partially overlaps with ORF2 but has a different reading frame (Rodriguez-Frias et al., 2012). It has been suggested that ORF3 might be an accessory protein of intracellular expression (Perez-Gracia et al., 2014) and that it probably affects the host response (Holla et al., 2013; Rodriguez-Frias et al., 2012). Its D1 domain is

Table 2
ORF1 polymorphic positions in reference sequences (PP%) and mutations in IC sequence.

Protein	Methyltransferase (AA 56–240)	Y domain (AA 216–433)	Polyproline region (variable)	Protease (AA 433–592)	X domain (AA 795–968)	Helicase (AA 970-1202)	RNA polymerase (AA 1217–1703)
Summarized function	mRNA ribosome binding, innate immunity (Holla et al., 2013)	Unknown	Adaptation (Holla et al., 2013) and immune response modulation (Karpe and Meng, 2012)	Protease (Suppiah et al., 2011), deubiquitination and cellular immunity (Karpe and Meng, 2012)	Unknown	Helicase (Holla et al., 2013)	Replication (Ahmad et al., 2011)
Length (AA)	184	218	Variable (1)	160	174	233	487
(PP%)	5	0.40		19	22	6	7
IC sequence mutations	No mutations	No mutations	No insertions. No deletions.	H495C	T938 M, T856 V, S898A	S1014N, S975T, Q1133 K	D1235G, Q1242R, S1454T, V1480I, I1502 V, K1511R, G1373 V, E1442D, V1693 M

(1) 110 AA in most of the sequences ($n = 43$); 114 AA ($n = 2$); 131 AA ($n = 3$); 172 AA ($n = 3$);107 AA ($n = 7$); 108 AA ($n = 3$); 109 AA ($n = 1$).

Table 3
ORF2 and ORF3 polymorphic positions in reference sequences (PP%) and mutations in IC sequence.

	ORF2 protein				ORF3 protein			
	S domain (N-terminal) (AA 1–113)	S domain (C-terminal) (AA 114–317)	M domain (AA 318–451)	P domain (AA 452–660)	D1 domain	D2 domain	P1 domain	P2 domain
Summarized function	Capsid (Xing et al., 2010)		Connecting S and P domains (Xing et al., 2010)	Neutralizing antibodies binding and cellular receptor (Holla et al., 2013)	Cell survival signal (Zafrullah et al., 1997)	Innate immune response Ratra, (Kar-Roy et al., 2008)	ORF2 interaction (ser71) (Ahmad et al., 2011)	Virus release (Ahmad et al., 2011)
Length (AA)	113	203	133	208	122			
(PP%)	22	1	0.70	6	0	16	35	25
IC sequence mutations	F10L, S26T, G36S, S70P, A105V, I113V	No mutations	T426A	No mutations	No mutations	A36V	No mutations	P110S (included in PXXP motif)

fully conserved in our analysis, in accordance with its function as a possible cell survival signal (Kar-Roy et al., 2004). The D2, P1 and P2 domains, which are related to the inhibition of innate host response, inflammatory response attenuation (Taneja et al., 2009), ORF2 interaction and virus release, harbor many polymorphic AAs. The previously described PMSP motif of the P1 proline-rich region (Ahmad et al., 2011) occurs in the PMS(F/Y) reference sequences, but not in any case of P80, suggesting that a mistake has been made in classifying residues into this motif. Amino acid S79, whose phosphorylation is determinant for the ORF3-ORF2 interaction (Tyagi et al., 2002), is fully conserved in all of them.

The analysis of the IC sequence indicates that the sequence from feces was fully representative of that from serum, making molecular analysis more feasible when serum samples are scarce. Methyltransferase and the Y domain showed no mutations with respect to the consensus. In contrast, PCP, the X domain, helicase and RdRp had 1, 3, 3 and 9 AA substitutions, respectively. Mutations accumulate predominantly in the intermediate portion of the RdRp, presumably being located outside the protein's active site, although the previously described III and IV motifs (Koonin, 1991) are affected. Conversely, the 3′ region and the GDD functional motif described as keys for replication (Agrawal et al., 2001) is fully conserved. The data provided by (Lhomme et al., 2014a) suggest that the proline-rich region has an important role in the persistence of the infection. The studied IC sequence harbors neither inser-

tion nor deletions in this genome region, which may be to do with the resolution of the infection. Future research on immunocompromised patients becoming persistently infected would provide data on the molecular basis of HEV persistence. ORF2 and ORF3 domains present no or single AA substitutions, with the exception of the N-terminal portion of S domain, which accumulates six mutations. We hypothesize that these differences in the capsid protein might be related to morphological or functional features, but they must surely represent phylogenetic differences relative to the reference sequences in terms of sub-genotype. Finally, we found a mutation in the IC sequence of the PXXP motifs of the P2 region associated with virus release (Pawson, 1995). Future directed mutation analysis might investigate the effect of this specific substitution on virus release.

5. Conclusions

We conclude that the designed tool allows the molecular analysis of the complete HEV coding genome sequences from clinical samples. Our findings are consistent with some previously described features of the genome, although the discrepancies highlight the need for further functional studies. The value of this type of analysis will become clear once more samples have been completely sequenced and examined in conjunction with their

associated clinical data and with the accumulated results from different types of infection.

References

Agrawal, S., Gupta, D., Panda, S.K., 2001. The 3′ end of hepatitis E virus. Virology 282, 87–101 (HEV) genome binds specifically to the viral RNA-dependent RNA polymerase (RdRp).

Ahmad, I., Holla, R.P., Jameel, S., 2011. Molecular virology of hepatitis E virus. Virus Res. 161, 47–58.

Avellón, A., P.C., Morago, L., Garcia-Galera, M.C., Echevarria, J.M., 2015. Manuscript in preparation.

Chandra, V., Taneja, S., Kalia, M., Jameel, S., 2008. Molecular biology and pathogenesis of hepatitis E virus. J. Biosci. 33, 451–464.

Dalton, H.R., Keane, F.E., Bendall, R., Mathew, J., Ijaz, S., 2011. Treatment of chronic hepatitis E in a patient with HIV infection. Ann. Intern. Med. 155, 479–480.

Fujiwara, S., Yokokawa, Y., Morino, K., Hayasaka, K., Kawabata, M., Shimizu, T., 2014. Chronic hepatitis E: a review of the literature. J. Viral Hepat. 21, 78–89.

Holla, R.P., Ahmad, I., Ahmad, Z., Jameel, S., 2013. Molecular virology of hepatitis E virus. Semin. Liver Dis. 33, 3–14.

Huzly, D., Umhau, M., Bettinger, D., Cathomen, T., Emmerich, F., Hasselblatt, P., Hengel, H., Herzog, R., Kappert, O., Maassen, S., Schorb, E., Schulz-Huotari, C., Thimme, R., Unmussig, R., Wenzel, J.J., Panning, M., 2014. AT Transfusion-transmitted hepatitis E in Germany, 2013. Euro Surveill. 19.

Kamar, N., Mansuy, J.M., Cointault, O., Selves, J., Abravanel, F., Danjoux, M., Otal, P., Esposito, L., Durand, D., Izopet, J., Rostaing, L., 2008. Hepatitis E virus-related cirrhosis in kidney- and kidney-pancreas-transplant recipients. Am. J. Transplant. 8, 1744–1748.

Kar-Roy, A., Korkaya, H., Oberoi, R., Lal, S.K., Jameel, S., 2004. The hepatitis E virus open reading frame 3 protein activates ERK through binding and inhibition of the MAPK phosphatase. J. Biol Chem. 279, 28345–28357.

Koning, L., Pas, S.D., de Man, R.A., Balk, A.H., de Knegt, R.J., ten Kate, F.J., Osterhaus, A.D., van der Eijk, A.A., 2013. Clinical implications of chronic hepatitis E virus infection in heart transplant recipients. J. Heart Lung Transplant., 78–85.

Koonin, E.V., 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J. Gen. Virol. 72 (Pt. 9), 2197–2206.

Koonin, E.V., Gorbalenya, A.E., Purdy, M.A., Rozanov, M.N., Reyes, G.R., Bradley, D.W., 1992. 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. U. S. A. 89, 8259–8263.

Lhomme, S., Abravanel, F., Dubois, M., Sandres-Saune, K., Mansuy, J.M., Rostaing, L., Kamar, N., Izopet, J., 2014a. Characterization of the polyproline region of the hepatitis E virus in immunocompromised patients. J. Virol. 88 (October (20)), 12017–12025.

Lhomme, S., Garrouste, C., Kamar, N., Saune, K., Abravanel, F., Mansuy, J.M., Dubois, M., Rostaing, L., Izopet, J., 2014b. Influence of polyproline region and macro domain genetic heterogeneity on HEV persistence in immunocompromised patients. J. Infect. Dis. 209, 300–303.

Li, S., Tang, X., Seetharaman, J., Yang, C., Gu, Y., Zhang, J., Du, H., Shih, J.W., Hew, C.L., Sivaraman, J., Xia, N., 2009. Dimerization of hepatitis E virus capsid protein E2s domain is essential for virus-host interaction. PLoS Pathog. 5, e1000537.

Mateos Lindemann, M.L., Morales, J.G., Fernandez-Barredo, S., Dominguez, M.R., Garcia de la Hoz, F., Halfon, P., Perez Gracia, M.T., 2010. Fulminant hepatitis E in a woman taking oral contraceptive medication. Am. J. Trop. Med. Hyg. 82, 12–15.

Purdy, Michael A., Tam, Albert W., Huang, Chiao-Chain, Yarbough, Patrice O., Reyes, G.R., 1993. Hepatitis E virus: a non-enveloped member of the 'alpha-like' RNA virus supergroup? Semin. Virol. 4, 319–326.

Moal, V., Legris, T., Burtey, S., Morange, S., Purgus, R., Dussol, B., Garcia, S., Motte, A., Gerolami, R., Berland, Y., Colson, P., 2013. Infection with hepatitis E virus in kidney transplant recipients in southeastern France. J. Med. Virol. 85, 462–471.

Passos, A.M., Heringer, T.P., Medina-Pestana, J.O., Ferraz, M.L., Granato, C.F., 2013. First report and molecular characterization of hepatitis E virus infection in renal transplant recipients in Brazil. J. Med. Virol. 85, 615–619.

Pawson, T., 1995. Protein modules and signalling networks. Nature 373, 573–580.

Peralta, B., Mateu, E., Casas, M., de Deus, N., Martin, M., Pina, S., 2009. Genetic characterization of the complete coding regions of genotype 3 hepatitis E virus isolated from Spanish swine herds. Virus Res. 139, 111–116.

Perez-Gracia, M.T., Mateos Lindemann, M.L., Caridad Montalvo Villalba, M., 2013. Hepatitis E: current status. Rev. Med. Virol. 23, 384–398.

Perez-Gracia, M.T., Suay, B., Mateos-Lindemann, M.L., 2014. Hepatitis E: an emerging disease. Infect. Genet. Evol. 22, 40–59.

Pichlmair, A., Schulz, O., Tan, C.P., Naslund, T.I., Liljestrom, P., Weber, F., Reis e Sousa, C., 2006. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5′-phosphates. Science 314, 997–1001.

Price, J., 2014. An update on hepatitis B, D, and E viruses. Top. Antivir. Med. 21, 157–163.

Purdy, M.A., Lara, J., Khudyakov, Y.E., 2012. The hepatitis E virus polyproline region is involved in viral adaptation. PLoS One 7, e35974.

Reyes, G.R., Huang, C.C., Tam, A.W., Purdy, M.A., 1993. Molecular organization and replication of hepatitis E virus (HEV). Arch. Virol. 7, 15–25.

Rodriguez-Frias, F., Jardi, R., Buti, M., 2012. Hepatitis E: molecular virology, epidemiology and pathogenesis. Enferm. Infecc. Microbiol. Clin. 30, 624–634.

Song, Y.J., Park, W.J., Park, B.J., Lee, J.B., Park, S.Y., Song, C.S., Lee, N.H., Seo, K.H., Kang, Y.S., Choi, I.S., 2014. Hepatitis E virus infections in humans and animals. Clin. Exp. Vaccine Res. 3, 29–36.

Suppiah, S., Zhou, Y., Frey, T.K., 2011. Lack of processing of the expressed ORF1 gene product of hepatitis E virus. Virol. J. 8, 245.

Tanaka, T., Takahashi, M., Kusano, E., Okamoto, H., 2007. Development and evaluation of an efficient cell-culture system for Hepatitis E virus. J. Gen. Virol. 88, 903–911.

Taneja, S., Sen, S., Gupta, V.K., Aggarwal, R., Jameel, S., 2009. Plasma and urine biomarkers in acute viral hepatitis E. Proteome Sci. 7, 39.

Te, H.S., Drobeniuc, J., Kamili, S., Dong, C., Hart, J., Sharapov, U.M., 2013. Hepatitis E virus infection in a liver transplant recipient in the United States: a case report. Transplant. Proc. 45, 810–813.

Tyagi, S., Korkaya, H., Zafrullah, M., Jameel, S., Lal, S.K., 2002. The phosphorylated form of the ORF3 protein of hepatitis E virus interacts with its non-glycosylated form of the major capsid protein, ORF2. J. Biol. Chem. 277, 22759–22767.

Versluis, J., Pas, S.D., Agteresch, H.J., de Man, R.A., Maaskant, J., Schipper, M.E., Osterhaus, A.D., Cornelissen, J.J., van der Eijk, A.A., 2013. Hepatitis E virus: an underestimated opportunistic pathogen in recipients of allogeneic hematopoietic stem cell transplantation. Blood 122, 1079–1086.

Xing, L., Li, T.C., Mayazaki, N., Simon, M.N., Wall, J.S., Moore, M., Wang, C.Y., Takeda, N., Wakita, T., Miyamura, T., Cheng, R.H., 2010. Structure of hepatitis E virion-sized particle reveals an RNA-dependent viral assembly pathway. J. Biol. Chem. 285, 33175–33183.

Yarbough, P.O., Tam, A.W., Fry, K.E., Krawczynski, K., McCaustland, K.A., Bradley, D.W., Reyes, G.R., 1991. Hepatitis E virus: identification of type-common epitopes. J. Virol. 65, 5790–5797.

Zafrullah, M., Ozdener, M.H., Kumar, R., Panda, S.K., Jameel, S., 1999. Mutational analysis of glycosylation, membrane translocation, and cell surface expression of the hepatitis E virus ORF2 protein. J. virol. 73, 4074–4082.