



## OPEN ACCESS

## EDITED BY

Jue Liu,  
Yangzhou University,  
China

## REVIEWED BY

Gaëtan Ligat,  
Université Toulouse III Paul Sabatier, France  
Yi Xu,  
Nanjing Agricultural University,  
China

## \*CORRESPONDENCE

Ana Avellon  
aavellon@isciii.es

## SPECIALTY SECTION

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 04 August 2022

ACCEPTED 20 September 2022

PUBLISHED 06 October 2022

## CITATION

Muñoz-Chimeno M, Rodriguez-Paredes V,  
García-Lugo MA and Avellon A (2022)  
Hepatitis E genotype 3 genome: A  
comprehensive analysis of entropy, motif  
conservation, relevant mutations, and  
clade-associated polymorphisms.  
*Front. Microbiol.* 13:1011662.  
doi: 10.3389/fmicb.2022.1011662

## COPYRIGHT

© 2022 Muñoz-Chimeno, Rodriguez-Paredes, García-Lugo and Avellon. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Hepatitis E genotype 3 genome: A comprehensive analysis of entropy, motif conservation, relevant mutations, and clade-associated polymorphisms

Milagros Muñoz-Chimeno<sup>1,2</sup>, Vanessa Rodriguez-Paredes<sup>1</sup>,  
Maira Alejandra García-Lugo<sup>1</sup> and Ana Avellon<sup>1,3\*</sup>

<sup>1</sup>Hepatitis Unit, National Center of Microbiology, Carlos III Institute of Health, Madrid, Spain, <sup>2</sup>Alcalá de Henares University, Madrid, Spain, <sup>3</sup>CIBERESP Epidemiology and Public Health, Madrid, Spain

Hepatitis E virus genotype 3 (HEV-3) is an EU/EEA emergent zoonosis. HEV-3 clades/subtypes have been described. Its genome contains ORF1, which encodes nonstructural proteins for virus replication, ORF2, the capsid protein, and ORF3, a multifunctional protein involved in virion pathogenesis. The study aims with respect to HEV-3 are to: (1) calculate genome entropy (excluding hypervariable region); (2) analyze the described motifs/mutations; (3) characterize clade/subtype genome polymorphisms. Seven hundred and five sequences from the GenBank database were used. The highest entropies were identified in zoonotic genotypes (HEV-3 and HEV-4) with respect to HEV-1 in X domain, RdRp, ORF2, and ORF3. There were statistically significant differences in the entropy between proteins, protease and ORF3 being the most variable and Y domain being the most conserved. Methyltransferase and Y domain motifs were completely conserved. By contrast, essential protease H581 residue and catalytic dyad exhibited amino acid changes in 1.8% and 0.4% of sequences, respectively. Several X domain amino acids were associated with clades. We found sequences with mutations in all helicase motifs except number IV. Helicase mutations related to increased virulence and/or fulminant hepatitis were frequent, the 1,110 residue being a typical HEV-3e and HEV-3f-A2 polymorphism. RdRp motifs III, V, VII also had high mutation rates. Motif III included residues that are polymorphisms of HEV-3e (F1449) and HEV-3m (D1451). RdRp ribavirin resistance mutations were frequent, mainly 1479I (67.4, 100% in HEV-3efglmk) and 1634R/K (10.0%, almost 100% in HEV-3e). With respect to ORF2, 19/27 neutralization epitopes had mutations. The S80 residue in ORF3 presented mutations in 3.5% of cases. Amino acids in the ORF3-PSAP motif had high substitution rates, being more frequent in the first PSAP (44.8%) than in the second (1.5%). This is the first comprehensive analysis of the HEV-3 genome, aimed at improving our knowledge of the genome, and establishing the basis for future genotype-to-phenotype analysis, given that viral features associated with severity have not been explored in depth. Our results demonstrate there are important genetic differences in the studied genomes that sometimes affect significant viral structures, and constitute

clade/subtype polymorphisms that may affect the clinical course or treatment efficacy.

#### KEYWORDS

hepatitis E virus, HEV, domains, polymorphisms, entropy, genome, genotypes, motif

## Introduction

Hepatitis E virus (HEV) is the most common cause of acute viral hepatitis worldwide (Aslan and Balaban, 2020), and is also responsible for severe chronic infections in immunocompromised patients. The number of EU/EEA-acquired infections has been increasing over the last decade (Adlhoch et al., 2016). HEV is a member of the *Hepeviridae* family and the *Orthohepevirus* genus, whose group A includes zoonotic viruses that infect humans, pigs, rabbits, wild boars and camels, among others (Purdy et al., 2017). Eight genotypes have been described so far, five of which can infect humans (HEV 1, 2, 3, 4, and 7; Smith et al., 2016). HEV genotype 3 (HEV-3) is a zoonosis. Pigs are its most important reservoir and it is transmitted through the consumption of raw or undercooked meat (Garcia et al., 2019). According to Smith et al. (2020), HEV-3 is further divided into subtypes 3a, 3b, 3c, 3e, 3f, 3g, 3h, 3i, 3j, 3k, 3l, 3m and some unassigned sequences. Subtypes are grouped in two major clades: HEV-3efg and HEV-3abjkchilm (Smith et al., 2016). The latter clade is further divided into HEV-3abjk and HEV-3chilm. A recent study divided HEV-3f into three clusters (HEV-3f-A1, HEV-3f-A2 and HEV-3f-B), and HEV-3m into HEV-3m-A and HEV-3m-B, based on p-distance and phylogenetic analysis (Munoz-Chimeno et al., 2022).

The HEV genome is a 7.2-kb single-stranded positive-sense RNA molecule containing three partially overlapping open reading frames (ORF1, ORF2, and ORF3; Tam et al., 1991). HEV-1 was recently shown to have an additional reading frame (ORF4) that overlaps with ORF1 and is required to ensure correct HEV RNA polymerase function (Nair et al., 2016). In 1992, computational analysis of the non-structural polyprotein identified seven putative domains [methyltransferase, Y domain, putative papain-like cysteine protease, proline-rich hinge domain, X domain, putative RNA helicase and putative RNA polymerase (Koonin et al., 1992)].

HEV methyltransferase (MTase) is characterized by 5'-mRNA capping activity through guanylyltransferase and guanine-7-methyltransferase activities (Kabrane-Lazizi et al., 1999; Magden et al., 2001). It also presents the four highly conserved sequence motifs described in the capping proteins of positive-strand RNA viruses (Rozanov et al., 1992). The A105H mutation in Mtase was associated with viremia decrease (Borkakoti et al., 2017) and F179S was associated with fulminant hepatitis (Mishra et al., 2013). The non-structural Y domain was initially found approximately 200 amino acid residues downstream of the MTase domain (Koonin et al., 1992). It contains critical conserved

residues that have been shown to play a crucial role in virus replication (Parvez, 2017; Cao et al., 2018) and adaptation (Shafat et al., 2021). HEV protease is a putative papain-like cysteine protease (PCP) and although its functions are still under debate, molecular analysis of mutant replicons identified highly conserved cysteines and histidines that are essential for replication (Parvez, 2013), catalytic dyad (Parvez and Khan, 2014) and a Zn<sup>2+</sup> binding site that is crucial for viral activity (Saraswat et al., 2019). The HEV X domain was classified in 2003 as the ADP-ribose-1"-monophosphatase of macrodomain protein family (Allen et al., 2003) and includes highly conserved residues that corroborate this classification (Parvez, 2015). In addition, C-terminal region presents residues whose interaction suggests a relevant role in the viral cycle (Anang et al., 2016), as well as Mg<sup>2+</sup> and Zn<sup>2+</sup> binding sites (Vikram and Kumar, 2018). The HEV X domain may play a role in viral replication and/or translation (Neuvonen and Ahola, 2009) and has been identified as a putative IFN antagonist *in vitro* (Nan et al., 2014). The putative RNA helicase (Hel) contains seven conserved motifs (I, Ia, II, III, IV, V, and VI) of the helicase's superfamily 1 (Gorbalenya et al., 1989; Hall and Matson, 1999). Hel mutations have been associated with virus virulence and disease severity (Inoue et al., 2009; Takahashi et al., 2009; Devhare et al., 2014; Cao et al., 2018). RdRp, which is phylogenetically classified in supergroup III, catalyzes the RNA viral synthesis at several levels (Oechslin et al., 2022) and contains a highly conserved GDD motif, which constitutes a catalytic triad associated with the replicative activity (Koonin, 1991; Rehman et al., 2008). Several mutations in the RdRp region have been associated with adverse clinical outcomes (Debing et al., 2014, 2016; Borkakoti et al., 2016, 2017; Todt et al., 2016). ORF2, which contains three domains, designated S, M, and P, was initially considered to encode only the capsid protein (Robinson et al., 1998; Graff et al., 2008), but in recent years has been shown to present itself in different forms with multiple functions rather than just acting as a viral capsid (Ankavay et al., 2019). Several mutations are associated with reduced replication and infectivity (Cordoba et al., 2011). Two initiation codons have recently been described (Yin et al., 2018). ORF2-encoded protein is also responsible for the humoral immune response and therefore harbors antigenic properties (Khudyakov et al., 1994; Zhou et al., 2005) and neutralization epitopes (Gu et al., 2015; Ikram et al., 2018). ORF3, the smallest ORF, partially overlaps with the N-terminus ORF2 and is translated from a different reading frame. ORF3 is essential for HEV infection and is required for viral particle release (Graff et al., 2005; Yamada et al., 2009).

Moreover, a reported motif within ORF3 protein has been shown to be required for membrane-associated HEV particle formation (Nagashima et al., 2011). Currently, HEV-ORF3 is thought to form an ion channel that is required for virion particle release from cells during infection (Ding et al., 2017). Similarly to ORF2, ORF3 has three distinct initiation codons (Graff et al., 2006).

This study aims to: (1) measure HEV-3, HEV-1 and HEV-4 genome entropy that has not yet been analyzed; (2) analyze conservation of HEV-3 genome functional motifs and mutations; and (3) identify and characterize HEV-3 subtype polymorphisms throughout the coding genome. The hypervariable region (HVR) was not analyzed in this study as this work has already been reported (Munoz-Chimeno et al., 2020).

## Materials and methods

### Sequence collection for genome analysis and sub-genotype assignment

HEV-3 genomes included in the analysis were retrieved from GenBank, 74 of which had previously been obtained in our laboratory (MZ289076-MZ289149; Muñoz-Chimeno et al., 2022) and 439 were additional complete genomes retrieved from GenBank database in October 2021 (detailed in Supplementary material). In addition, 60 HEV-1 and 92 HEV-4 sequences from GenBank were used for entropy analysis (see Supplementary material). Reference sequences proposed by Smith et al. (2020) and Muñoz-Chimeno et al. (2022) were used.

Sub-genotypes were checked by phylogenetic analysis as described elsewhere (Munoz-Chimeno et al., 2022).

### Alignment, translation and motif/position identification

Sequences were aligned using MAFFT-based alignment v.7 [MAFFT alignment and NJ/UPGMA phylogeny (cbrj.jp)]. Protein translation was carried out with the MEGA 7.0 software package.<sup>1</sup> The hypervariable region was then removed from the alignment. The beginning and end of each ORF and protein were identified according to Koonin et al. and individual protein alignments were generated. Amino acids were numbered according to the position of each residue in each ORF (ORF numbering) according to the sequence with GenBank accession number KU513561. Additionally, in the case of ORF1, protein numbering (from 1 to the end of each protein) is set out in the text and Supplementary material. Previously reported relevant motifs or positions were localized by searching for their sequences in the alignments.

<sup>1</sup> <https://www.megasoftware.net/>

## Entropy analysis of the genome of HEV proteins

Shannon entropy, calculated with the Antigenic Variability ANalyzer (AVANA) tool, was considered to be a measure of the variability of HEV protein sequences. Only sequences containing a valid amino acid at each position were used for the entropy analysis. Entropy was calculated by position first in the HEV-1, HEV-3 and HEV-4 genotype alignment, and second in the HEV-3 main clades (HEV-3abjk, HEV-3chilm and HEV-3efg). Levels of entropy were categorized as low (<0.03), intermediate (0.03–0.06) or high (>0.06).

## Statistics

Mean entropy in each protein and in each clade and protein was compared by ANOVA and *post hoc* Scheffe tests in IBM SPSS Statistics v.25.0. The threshold of significance was set at  $p < 0.05$ .

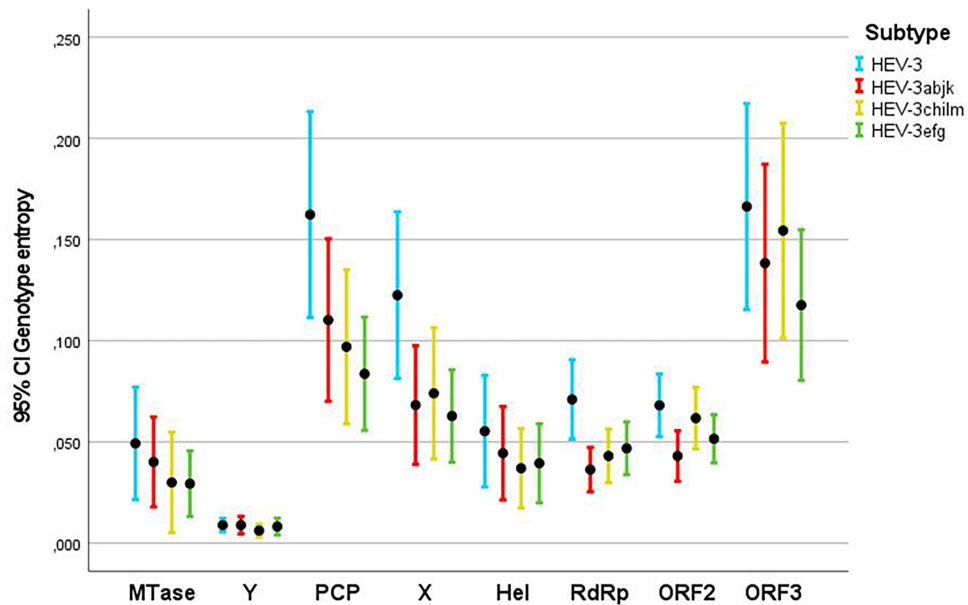
## Identification of polymorphisms associated with clades and sub-genotypes

To identify the polymorphisms associated with clades and subtypes, the percentages of each amino acid in each position by clade and subtype were calculated. To be considered polymorphisms and subtype/clade characteristic residues, the percentage of the specific amino acid should be at least 90% that in the subtype/clade.

## Results

### HEV genome entropy

Average entropies by protein/ORFs (excluding HVR) in the HEV-3 genome and by clade are shown in Figures 1, 2. Several statistically significant differences were observed. The PCP and ORF3 showed higher average entropy (0.162 and 0.165, respectively) than MTase, the Y domain, Hel, RdRp and ORF2 ( $p < 0.05$ ). The Y domain had the lowest entropy in all clades (0.009) with respect to the PCP, X domain and ORF3 (0.162, 0.123, and 0.165, respectively). There were no statistically significant differences among the HEV-3 clades. The entropy at several positions along the HEV-3 genome was  $\geq 0.6$  (Table 1). PCP and ORF3 were proteins with a higher percentage of high-entropy positions (10.7% and 12.3% of total protein, respectively). ORF2 was the protein with the highest number of high-entropy positions ( $n = 24$ ). By contrast, the Y domain did not have any high-entropy positions. The most variable positions in the X domain were 906 and 938. High-entropy positions were noted in



	MTase	Y domain	PCP	X domain	Hel	RdRp	ORF2	ORF3
HEV-3	0.045	0.009	0.162	0.123	0.055	0.071	0.068	0.165
HEV-3abjk	0.036	0.009	0.110	0.068	0.045	0.036	0.043	0.137
HEV-3chilm	0.027	0.006	0.097	0.074	0.037	0.043	0.062	0.153
HEV-3efg	0.028	0.008	0.084	0.063	0.040	0.047	0.052	0.117

**FIGURE 1**  
 Error bar graph showing the mean entropy and 95% confidence interval (CI) for each protein in HEV-3, HEV-3abjk, HEV-3chilm, and HEV-3efg. The table below shows the average entropies of each protein. PCP and ORF3 exhibited statistically significant differences compared with MTase, Y domain, Hel, RdRp, and ORF2.

all HEV-3 clades, the HEV-3chilm clade containing more of these than the other clades.

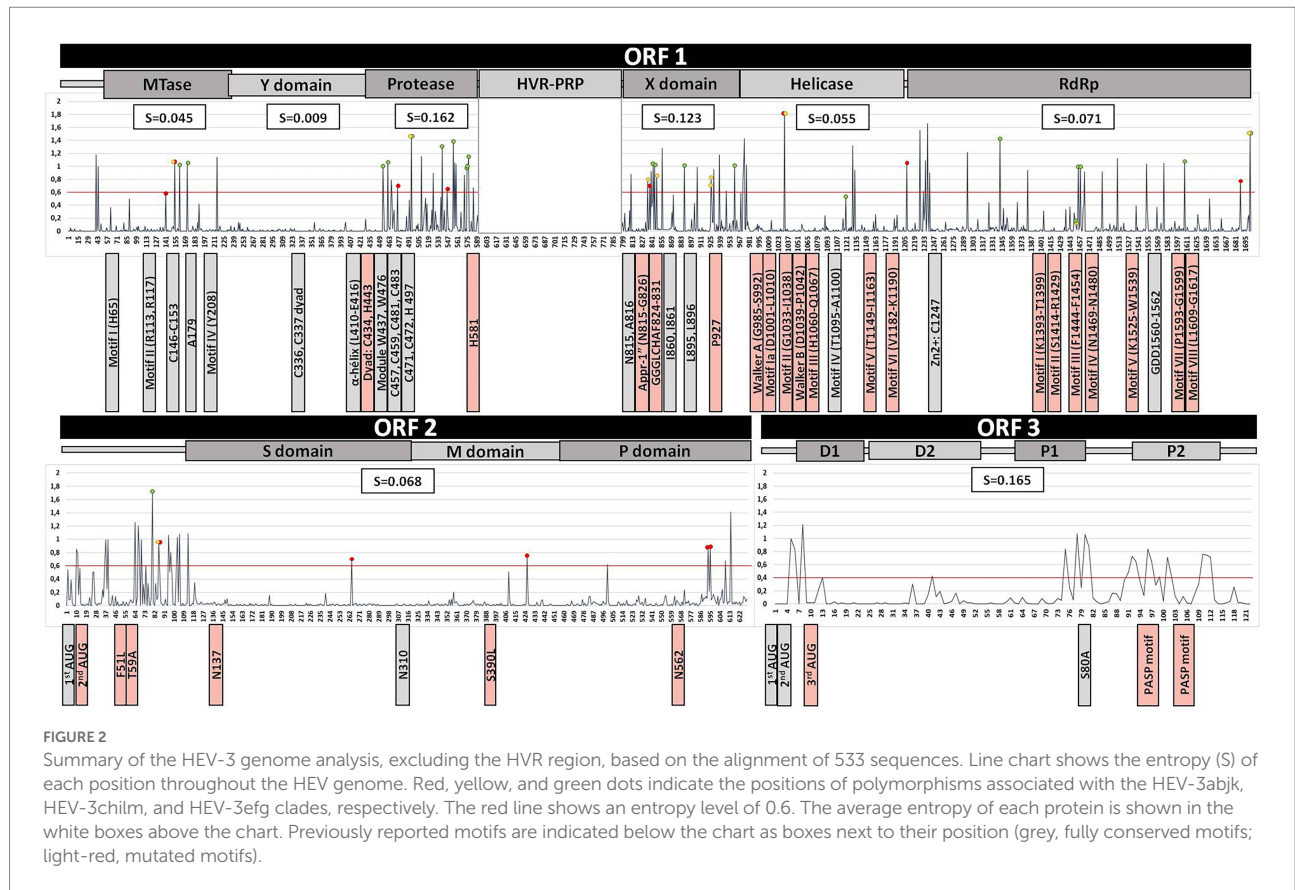
Entropy was also assessed by genotype (HEV-1, HEV-3, and HEV-4) and protein (Figure 3). The highest entropies were identified in genotypes HEV-3 and HEV-4 (zoonotic genotypes) relative to HEV-1 (main non-zoonotic genotype) in the X domain, RdRp, ORF2 and ORF3 ( $p < 0.05$ ).

### Analysis of previously reported HEV genome motifs and mutations

Results of the ORF1 analysis of the HEV-3 genome are summarized in Table 2 and Figure 2. Previously reported motifs in the **Mtase** and **Y domain** are fully conserved, while an Mtase position known to be associated with fulminant hepatitis in HEV-1 (F179) was found to differ in the HEV-3 sequences analyzed, the wild type being 179A. Neither the palmitoylation site nor the  $\alpha$ -helix of the **Y domain** had substitutions. **PCP** motifs associated with protease activity and structural integrity

(C and W residues, respectively) were all conserved, with the exception of C434 (0.2% C434S). Notably, mutations of Y, Q and R were observed in H581 (1.9%) and of H in Y443 (0.2%). Reviewing the motifs and known mutations of the **X domain** revealed a change in G826D in 0.2% of sequences, and mutations in the metal binding sites G826D, L827I, and H829LR (0.9%) and P927L (14.1%), which is characteristic of the HEV-3c subtype. Analysis of **Hel** revealed mutations in every described motif except motif IV. The others exhibited changes ranging from 0.2% to 6.4%. Two sequences had at least two mutations in the Walker A motif. The two mutations reported to be associated with fulminant hepatitis in HEV-1 (L1110F and V1120I) were found in 0% and 51.9% of cases, respectively. Nevertheless, L1110MQRSV was frequently found (9.9%), 1110M being characteristic of the HEV-3e and HEV-3f-A2 clades. The mutation related to increased virulence in HEV-3 (V1208A) was present in 2.4% of the sequences studied and found to be characteristic of HEV-3b. The zinc-binding motif and replication catalytic triad of **RdRp** were fully conserved. By contrast, substitutions were found to affect motifs I (0.2%), II





**FIGURE 2**  
Summary of the HEV-3 genome analysis, excluding the HVR region, based on the alignment of 533 sequences. Line chart shows the entropy (S) of each position throughout the HEV genome. Red, yellow, and green dots indicate the positions of polymorphisms associated with the HEV-3abjk, HEV-3chilm, and HEV-3efg clades, respectively. The red line shows an entropy level of 0.6. The average entropy of each protein is shown in the white boxes above the chart. Previously reported motifs are indicated below the chart as boxes next to their position (grey, fully conserved motifs; light-red, mutated motifs).

(0.8%), III (13.3%), IV (0.9%), V (7.9%), VII (2.8%) and VIII (0.9%). Of the seven previously described ribavirin resistance mutations, the most common were V1479I (67.2%) and G1634RK (10.3%).

Table 3 and Figure 2 summarize the ORF2 and ORF3 analysis of the HEV-3 genome. The ORF2 second start codon had 0.2% of substitutions, while the first codon was fully conserved. Glycosylation sites S137 and T562 were mutated in 0.6 and 2.1%, respectively. Residues whose changes were related to reduced replication and infectivity were mutated in 0.2% (F51L), 0.4% (T59A) and 0.2% (S390L) of analyzed sequences. Of the previously described HEV-3 neutralization epitopes, the P491 epitope did not appear to be mutated in any of the sequences. The remaining epitopes detailed in Table 3 were mutated (0.2%–1.7%). In the case of the epitopes described in HEV-1 and HEV-4, whose mutation frequencies are detailed in Table 3, the V606A substitution was present in 96.2% and L613T was represented in 99.1% of HEV-3 sequences. The first and second start codons of ORF3 were both fully conserved, and a mutation in the third codon (M10A, 0.2%) was also noted. Position S80, which has been associated with the ORF2–ORF3 interaction, had S80LPF in 3.6% of its sequences. Finally, PSAP motifs involved in replication had a high substitution rate (mainly the first one motif) ranging from 0.2% to 25.7% in the residues involved. One sequence harbored mutations in both PSAP motifs.

### Clade/subtype-associated polymorphisms

Some amino acid polymorphisms were characteristic of different clades/subtypes, by which we mean that up to 90% of sequences of the clade/subtype harbor a particular polymorphism (Supplementary Table 1; Tables 2, 3). In MTase, position 1611 was characteristic of HEV-3efg, while 172S was specific to subtype HEV-3f, 141S of HEV-3k and 154S of HEV-3k and HEV-3c. In PCP 454A, 461S, 495C, 559G, 575S, 576I, and 577T are characteristic of HEV-3efg/ef, and 475R characterizes HEV-3abjk. The 495C, 546A, and 539D residues were characteristic of HEV-3m, HEV-3b and HEV-3f-A1, respectively. No clade or subtype polymorphisms were described in the Y domain. In the X domain, 838A was associated with HEV-3abjk; 843D, 888I, and 960S were associated with HEV-3efg/ef; and 835Y, 848D and 926D were linked to HEV-3chilm/chil. 845V is characteristic of HEV-3f subtype. In Hel, position 1,032 featured a different amino acid in each of the HEV-3 subtypes (Supplementary Table 1). In the case of RdRp, 1342G, 1455S, 1458I, 1608A, and 1702I are characteristic of the HEV-3efg clade, as 1688I is of HEV-3abjk. Furthermore, 1449F (motif I) was found in HEV-3e and 1451D (motif III), and 1702M was present in HEV-3m. HEV-3abjk presented several characteristic mutations in ORF2: 86A, 264T, 426T, 593T, and 595I. 86T was found in HEV-3hl subtypes and 80AV was present

TABLE 1 Amino acid positions with entropy  $\geq 0.6$  throughout the genome.

Protein/ORF	HEV-3 amino acid positions with entropy $\geq 0.6$	N entropy $\geq 0.6$	Total AA in the protein	% entropy $\geq 0.6$
Mtase	141, 154, 161, 172, 215	5	185	2.7
Y domain	–	0	218	0.0
PCP	454, 461, 466, 475, 495, 509, 526, 539, 546, 555, 557, 559, 571, 575, 576, 577, 584	17	159	10.7
X domain	811, 835, 838, 840, 843, 845, 848, 856, 888, 906, 926, 930, 938, 948, 960	15	172	8.7
Helicase	972, 973, 974, 977, 1,032, 1,130, 1,133	7	233	3.0
RdRp	1,227, 1,232, 1,235, 1,238, 1,241, 1,295, 1,342, 1,382, 1,455, 1,458, 1,464, 1,489, 1,511, 1,553, 1,578, 1,608, 1,688, 1702	18	487	3.7
ORF2	10, 11, 36, 37, 39, 64, 67, 68, 70, 74, 80, 86, 95, 97, 103, 105, 113, 264, 426, 500, 593, 595, 609, 614	24	629	3.8
ORF3	5, 6, 8, 75, 78, 80, 81, 92, 93, 96, 97, 101, 110, 111, 112	15	122	12.3

Numbering is set according to ORF.

in the HEV-3f subtype. **ORF3** contained no clade or subtype polymorphisms.

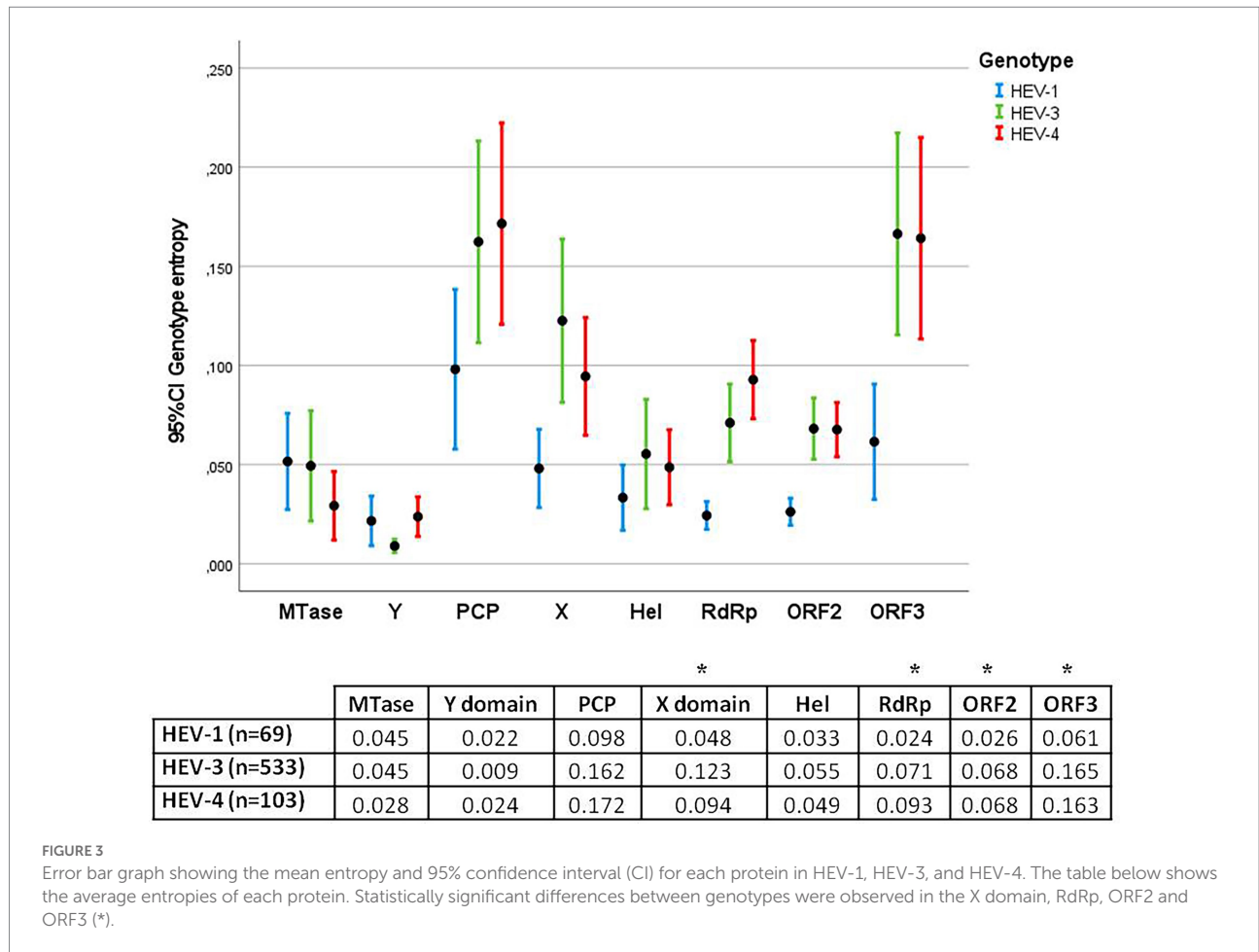
## Discussion

HEV is considered an emerging infection in Europe, and is known to be of great clinical relevance in immunocompromised patients. Disease severity depends on host immune status (von Felden et al., 2019) and it has been suggested that this might be associated with specific nucleotide substitutions (Inoue et al., 2009). In fact, HEV is composed of a wide range of virus genotypes and subtypes whose relationships between genetic characteristics and clinical manifestations are yet to be fully explored. The HEV genome is not completely understood, and the PCP, Y domain and HVR still await a clear functional assignment. In this context, we analyzed the HEV genome in depth.

First, we analyzed the variability of the HEV genome and, for the first time, examined the entropy and identified significant differences among genotypes; in short, we found the entropy of zoonotic genotypes to be higher than that of non-zoonotic genotypes in the X domain, RdRp, ORF2 and ORF3. Entropy reflects the diversity of circulating viruses and is somehow related to epidemiological behavior, including the emergence of new variants, as suggested by studies of SARS CoV-2 epidemic in India (Santoni et al., 2022) and mumps outbreaks in Spain (Gavilan et al., 2022). The greater diversity of circulating zoonotic viruses might be related to the more frequent introduction of variants due to the broader diversity of hosts or the wider geographical distribution of HEV-3 and HEV-4 genotypes. It is striking that no significant differences were found among the HEV-3 clades, suggesting that different HEV-3 clades might have similar epidemiological behavior. When analyzing entropy differences among the HEV genome proteins, excluding HVR, we found the Y domain to

be extremely highly conserved, suggesting that this protein has a crucial role in the viral cycle.

Second, we analyzed the previously described motifs and mutations in HEV-3 ORF1, of which the completely conserved Mtase and Y domain motifs (Rozanov et al., 1992; Cao et al., 2018) were of particular note. By contrast, the PCP essential H581 residue and catalytic dyad C434-H443 (Parvez, 2013) exhibited amino acid changes in 1.8% and 0.4% of sequences, respectively. This finding is in line with the still poorly defined functions of the PCP region. The X domain is present in the *Hepeviridae*, *Togaviridae* and *Coronaviridae* families, where it is considered a putative IFN antagonist that modulates the host immune response as well as being associated with viral pathogenesis and replication (Li et al., 2016). In addition, both HEV X domain and polyproline region heterogeneity were proposed as being associated with viral persistence (Lhomme et al., 2014). Interestingly, our results show that the HEV-3 X domain has several polymorphic amino acids that are associated with clades, an observation that might be consistent with the recently reported clade differences in clinical behavior (Schemmerer et al., 2022). Additionally, it has recently been suggested that the conserved macrodomain is a potential therapeutic target for coronavirus and alphavirus (Leung et al., 2022). The entropy of the X domain is among the highest in our study, so the treatment target of this genome region might not be an option for HEV-3. We found sequences with mutations in different residues of all Hel motifs except motif IV. Motifs I, II and VI mediate binding and hydrolysis of NTP, while III and VI are involved in coupling ATPase activity to helicase function (Hall and Matson, 1999). Also in the Hel, point mutations have been described in positions 1,110, 1,120, and 1,208 that are related to increased virulence and/or fulminant hepatitis (Takahashi et al., 2009; Devhare et al., 2014). These point mutations are frequently observed in the studied sequences (especially I<sup>1120</sup>, which occurs in 52.1% of sequences).



**FIGURE 3**  
 Error bar graph showing the mean entropy and 95% confidence interval (CI) for each protein in HEV-1, HEV-3, and HEV-4. The table below shows the average entropies of each protein. Statistically significant differences between genotypes were observed in the X domain, RdRp, ORF2 and ORF3 (\*).

Interestingly, the 1,110 residue constitutes a polymorphism typical of HEV-3e and the recently described HEV-3f-A2 (Muñoz-Chimeno et al., 2022), while the 1,208 residue is typical of HEV-3b. This finding suggests a possible association of the subtype with severity of the disease. Although information about the clinical course of cases is not available, most of the complete genomes available in the GenBank database presumably correspond to clinical cases. It would be interesting to carry out prospective studies with clinical information to assess whether the presence of these mutations is linked to the clinical course. The study of these mutations might be useful for patient management and prediction outcome. RdRp motifs III, V and VII also had high mutation rates. Motif III includes residues that are polymorphisms of HEV-3e (F1449) and HEV-3m (D1451). Ribavirin is currently used as first-line antiviral therapy to treat severe or chronic HEV infection (Raji et al., 2022). *In vitro* studies have identified associations between some RdRp point mutations and treatment failure (Debing et al., 2014, 2016; Borkakoti et al., 2016, 2017; Todt et al., 2016). These mutations are very frequent in the analyzed sequences, especially 1479I and 1634R/K, which are present in 67.4% and 10.0% of sequences, respectively. 1479I is present in 100% of HEV-3efg, HEV-3l, HEV-3m, and HEV-3k, while 1634R occurs in almost 100% of HEV-3e. This means that,

in the case of ribavirin treatment, it might be necessary to have subtype information in order to predict antiviral response, especially when a long-term treatment is needed.

Third, two initial codons have been described in HEV ORF2 that are of uncertain significance in the viral cycle (Yin et al., 2018). The second initial codon could give rise to a strategy of immune toleration similar to that of hepatitis B virus (HBV). HBV generates “e” protein using an alternative to the core protein initial codon, the action being related to a milder immune response and less cellular damage. Mutation of this initial codon or the generation of a stop codon causes a severe and rapidly evolving HBV infection (Malik et al., 2018). One of the studied sequences has a mutation in this alternative codon. It would be very interesting to examine the relationship between the presence of these mutations and the clinical course of the infection. Neutralization epitopes have been described for HEV-1, HEV-3, and HEV-4 (Gu et al., 2015; Ikram et al., 2018) in 27 residues of ORF2, of which 430, 433, 476, 477, 485, 491, 512, and 549 had no mutations in the HEV-3 sequences studied. The remaining 19 sequences had mutation rates between 0.2% and 99.1%. These results suggest that immune epitopes are highly variable and highlight the importance of developing a pan-genotypic vaccine (Nan et al., 2018).

**TABLE 2** Analysis of previously reported motifs and mutations of ORF1 of HEV-3 genome (references of motifs and mutation are mentioned in the Introduction and listed in full in [Supplementary material](#)).

	<b>Functional motif /mutation (Position according to ORF) (Associated with)</b>	<b>Residues (Position beginning each protein)</b>	<b>Substitutions sequences (n; %)</b>	<b>Clade/subtype polymorphism</b>
Mtnase	Motif I, II, IV (H <sup>65</sup> , DXXXXR <sup>113–117</sup> , Y <sup>208</sup> )		0	I <sup>161</sup> (3efg) S <sup>172</sup> (3f) S <sup>141</sup> (3k) S <sup>154</sup>
	C residues (C <sup>146</sup> -C <sup>153</sup> )		0	(3k) S/P <sup>154</sup> (3c)
	A27V; N29D (Increased viral load and severity)		0	
	R105H (Decreased viremia)		0	
	F179S (Fulminant hepatitis in HEV-1)	124	A179F (1; 0.2%)	
Y	Cysteine dyad (C <sup>336</sup> -C <sup>337</sup> ) (Palmitoylation site)		0	–
	α-helix (LYSWLFE <sup>410–416</sup> ) (Inhibit infectivity)		0	
PCP	CRC <sup>457–459</sup> CTC <sup>481–483</sup> (Zn <sup>2+</sup> binding. Loss of activity)		0	A <sup>454</sup> S <sup>461</sup> C <sup>495</sup> G <sup>559</sup> (3efg); D <sup>539</sup> (3f-)
	C <sup>471</sup> , C <sup>472</sup> (C471A, C472A suppressed replication)		0	A1) R <sup>475</sup> (3abjk) C <sup>495</sup> (3m); A <sup>546</sup>
	H <sup>497</sup> , H <sup>581</sup> (H497L, H581L suppressed replication)	H <sup>64</sup> , H <sup>148</sup>	YQR <sup>581</sup> (10; 1.9%)	(3b) S <sup>575</sup> I <sup>576</sup> T <sup>577</sup> (3ef)
	Catalytic dyad C <sup>434</sup> -H <sup>443</sup> (PCP function)	C <sup>1</sup> -H <sup>10</sup>	S <sup>34</sup> , Y <sup>443</sup> (2; 0.4%)	
X domain	W module (W <sup>437</sup> -W <sup>476</sup> ) (Structural integrity)		0	
	NxxxNxxxHxxGGG <sup>815–826</sup> (Appr-1 <sup>o</sup> -pase active site that formed the secondary structure)	19–30	D <sup>826</sup> (1; 0.2%)	Y <sup>835</sup> (3chil) A <sup>838</sup> (3abjk) D <sup>843</sup> S <sup>960</sup>
	(I <sup>860</sup> -I <sup>861</sup> /L <sup>895</sup> -L <sup>896</sup> ) (ORF3 interaction, life-cycle)		0	(3efg) V <sup>845</sup> (3f); L <sup>888</sup> (3ef) D <sup>848</sup> D <sup>926</sup>
	N <sup>815</sup> , A <sup>816</sup> (Metal binding sites)		0	(3chilm)
	GGGLCHAF <sup>824–831</sup> (Metal binding sites)	28–35	D <sup>826</sup> , I <sup>827</sup> , LR <sup>829</sup> (5; 0.9%)	
Helicase	P <sup>927</sup> (Metal binding sites)	131	L <sup>927</sup> (75; 14.1%) (3c)	
	Motif I (Walker A) (GVPGSGKS <sup>985–992</sup> ) (ATPase activity, K <sup>991</sup> mutation abolish it)	16–23	D <sup>988</sup> , Y <sup>989</sup> , E <sup>990</sup> , THE <sup>991</sup> , T <sup>992</sup> (3; 0.6%)	I <sup>1032</sup> (3k) L <sup>1032</sup> (3h) T <sup>1032</sup> (3m) A <sup>1032</sup>
	Motif Ia (DVVVVPTREL <sup>1001–1010</sup> ) (ATPase activity)	32–41	H <sup>1010</sup> (1; 0.2%)	(3i, 3ef) G <sup>1032</sup> (3g)
	Motif II &Walker B: GRRVVI <sup>1033–1038</sup> /DEAP <sup>1039–1042</sup> (D residue interacts with Mg <sup>2+</sup> , NTPase activity)	64–73	C <sup>1034</sup> , I <sup>1037</sup> (3; 0.6%)	
	Motif III (HLLGDPNQ <sup>1060–1067</sup> ) (ATPase activity)	91–98	L <sup>1060</sup> , H <sup>1067</sup> (2; 0.4%)	
	Motif IV (THRCPA <sup>1095–1100</sup> )		0	
	Motif V (TVHEAQGATFTETTTI <sup>1149–1163</sup> )	180–194	T <sup>1153</sup> , Y <sup>1158</sup> , DKV <sup>1160</sup> , V <sup>1163</sup> (34; 6.4%)	
	Motif VI (VALTRHTEK <sup>1182–1190</sup> ) (binding, hydrolysis)	213–221	G <sup>1189</sup> , R <sup>1190</sup> (17; 3.2%)	
	L1110F (Fulminant hepatitis in HEV-1)	151	MQRSV <sup>1110</sup> (53; 9.9%) (3e, 3f-A2)	
	V1120I (Fulminant hepatitis in HEV-1)	161	I <sup>1120</sup> (277; 51.9%) TS <sup>1120</sup> (41; 7.7%)	
RdRp	V1208A (Increased virulence in HEV-3)	239	A <sup>1208</sup> (13; 2.4%) (3b) IT <sup>1208</sup> (386; 71.1%)	
	C <sup>1247</sup> (Zn <sup>2+</sup> binding motif)		0	G <sup>1342</sup> S <sup>1455</sup> I <sup>1458</sup> A <sup>1608</sup> I <sup>1702</sup> (3efg) F <sup>1449</sup>
	Motif I (KDCNKFT <sup>1393–1399</sup> )	177–183	N <sup>1393</sup> (1; 0.2%)	(motif III) (3e) I <sup>1688</sup> (3abjk) D <sup>1451</sup>
	Motif II (SAWSKTFCALFGPWFR <sup>1414–1429</sup> )	198–213	R <sup>1416</sup> , L <sup>1420</sup> (4; 0.8%)	(motif III) (3m) M <sup>1702</sup> (3m)
	Motif III (FYGDAYEESVF <sup>1444–1454</sup> )	228–238	L <sup>1444</sup> , S <sup>1448</sup> , F <sup>1449</sup> , DGK <sup>1450</sup> , DG <sup>1451</sup> , T <sup>1452</sup> , M <sup>1453</sup> , L <sup>1454</sup> (71; 13.3%)	
	Motif IV (NDFSEFDSTQNN <sup>1469–1480</sup> )	253–264	S <sup>1479</sup> , D <sup>1480</sup> (5; 0.9%)	
	Motif V (KHSGEPTLLWNTVW <sup>1525–1539</sup> )	309–323	S <sup>1531</sup> , I <sup>1538</sup> (42; 7.9%)	
	GDD <sup>1560–1562</sup> (Replication catalytic triad)		0	
	Motif VII (PIGLYAG <sup>1593–1599</sup> )	377–383	S <sup>1593</sup> , MV <sup>1594</sup> , FH <sup>1597</sup> (15; 2.8%)	
	Motif VIII (LPDVVRFAG <sup>1609–1617</sup> )	393–401	D <sup>1611</sup> , I <sup>1612</sup> (5 0.9%)	
	Y1320H, K1383N, D1384G, K1398R, V1479I, Y1587E, G1634R/K (Ribavirin treatment failure)	114, 177, 178, 192, 273, 381, 428	FHN <sup>1320</sup> (3; 0.6%); N <sup>1383</sup> (1; 0.2%); I <sup>1479</sup> (358; 67.2%); FH <sup>1587</sup> (2; 0.4%); RK <sup>1634</sup> (55; 10.3%)	



**TABLE 3** Analysis of previously reported motifs, mutations and neutralization epitopes (NE) of ORF2 and ORF3 of HEV-3 genome (references of motifs and mutation are mentioned in the Introduction and listed in [Supplementary material](#)).

	<b>Functional motif/mutation (Associated to)</b>	<b>Substitutions sequences (N; %)</b>	<b>Clade/subtype polymorphism</b>
ORF2	First AUG (M <sup>1</sup> ) Capsid-associated ORF2	0	A <sup>86</sup> , T <sup>264</sup> , T <sup>426</sup> , T <sup>593</sup> , T <sup>595</sup> (3abjk) T <sup>86</sup>
	Second AUG (M <sup>16</sup> ) ORF secreted form	I <sup>16</sup> (1; 0.2%)	(3h, 3l) A/V <sup>80</sup> (3f)
	Glycosylation sites: NLS <sup>137</sup> NLT <sup>310</sup> NTT <sup>562</sup> (N562QDPY affects ORF2 dimerization and HEV infectivity)	DFK <sup>137</sup> (3; 0.6%) D <sup>562</sup> (11; 2.1%)	
	F51L (Decrease replication and infectivity)	L <sup>51</sup> (1; 0.2%)	
	T59A (Decrease replication and infectivity)	A <sup>59</sup> (2; 0.4%) N <sup>59</sup> (1; 0.2%)	
	S390L (Decrease replication and infectivity)	L <sup>390</sup> (1; 0.2%)	
	D430, L433, Y485, and R512 NE in HEV-1	0	
	E479	GK <sup>479</sup> (2; 0.4%)	
	D496	NY <sup>496</sup> (4; 0.8%)	
	I529	VT <sup>529</sup> (10; 1.9%)	
	K534	R <sup>534</sup> (1; 0.2%)	
	H577	R <sup>577</sup> (1; 0.2%)	
	R578	P <sup>578</sup> (1; 0.2%)	
	V606A NE in HEV-3	A <sup>606</sup> (513; 96.2%) GT <sup>606</sup> (2; 0.4%)	
	S487	PT <sup>487</sup> (2; 0.4%)	
	S488	P <sup>488</sup> (2; 0.4%)	
	T489	A <sup>489</sup> (1; 0.2%)	
	P491	0	
	D496	NY <sup>496</sup> (4; 0.8%)	
	Y561	F <sup>561</sup> (1; 0.2%)	
	T564	AINS <sup>564</sup> (4; 0.8%)	
	T585	N <sup>585</sup> (1; 0.2%)	
	T586	HP <sup>586</sup> (2; 0.4%)	
	P592	LJ <sup>592</sup> (7; 1.3%)	
	L613T NE in HEV-4	T <sup>613</sup> (537; 99.1%) AI <sup>613</sup> (3; 0.6%)	
	L477T (L476T)	0	
	E549 NE in HEV-1 and 4	0	
	K554	R <sup>554</sup> (12; 2.3%)	
	T497 NE in HEV-1 and 3	A <sup>497</sup> (2; 0.4%)	
	G591 NE in HEV-1, 3 and 4	AES <sup>591</sup> (9; 1.7%)	
ORF3	First AUG (M <sup>1</sup> )	0	–
	Second AUG (M <sup>2</sup> )	0	
	Third AUG (M <sup>10</sup> ) (True initial ORF3 protein)	A <sup>10</sup> (1; 0.2%)	
	S80A (S79A) (V66G-ORF2) (Assembly and ORF2/3 interaction)	LFP <sup>80</sup> (19; 3.6%)	
	PSAP motif (first) PSAP <sup>95–98</sup> (Mutations in both motifs decrease replication)	HLR <sup>95</sup> (8; 1.5%) LP <sup>96</sup> (137; 25.7%) VSG <sup>97</sup> (73; 13.7%) QHLR <sup>98</sup> (21; 3.9%)	
	PSAP motif (second) PSAP <sup>104–107</sup> (Mutations in both motifs decrease replication)	NRC <sup>105</sup> (7; 1.3%) V <sup>106</sup> (1; 0.2%)	

Finally, no mutation was found in the first and second codons of HEV ORF3, although one was found in one sequence in the third codon. Mutation in assembly interaction serine ORF2-ORF3 (S80) was detected in 3.5% of sequences to L, F and P. Although amino acid change described as affecting assembly was S80A due to the loss of phosphorylated S ([Nagashima et al., 2011](#)), none of

the amino acids found are known to be phosphorylated, so assembly could also be impaired. Amino acids in the PSAP motif had a high substitution rate, being more frequent in the first PSAP motif (44.8% sequences with substitutions) than in the second (1.5%). One sample featured a mutation in both PSAP motifs. PSAP motifs are required for transport machinery and promote

HEV envelopment and exit *via* the exome pathway. At least one of the two PSAP motifs is needed to form the membrane-associated HEV particles (Nagashima et al., 2011). According to our results, the second PSAP motif appears to be more important in this respect than the first.

This study, together with the HVR analysis previously reported (Munoz-Chimeno et al., 2020), constitute the first comprehensive genome analysis of HEV-3. The aim was to provide not only to improve our knowledge of the genome, but also to establish the basis for future genotype-to-phenotype analyses. HEV-3 is a virus with low pathogenicity in immunocompetent patients, but the viral features associated with its severity have not been explored in depth. Our results reveal important genetic differences in the studied genomes that sometimes affect significant viral structures, and that constitute clade or subtype polymorphisms that may affect the clinical course or treatment efficacy.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

The sequences obtained in this study were based on routine HEV testing, did not involve any additional sampling or tests, and irreversibly anonymized RNA extracts were used, so specific ethical approval was not required.

## Author contributions

MM-C performed the laboratory assays and sequence analysis and wrote the article. VR-P and MG-L helped carry out the

laboratory assays. AA directed and coordinated the study and the production of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

MM-C was funded by CIBERESP.

## Acknowledgments

The authors thank the Genomics Department of ISCIII.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1011662/full#supplementary-material>

## References

- Adlhoc, C., Avellon, A., Baylis, S. A., Ciccaglione, A. R., Couturier, E., de Sousa, R., et al. (2016). Hepatitis E virus: assessment of the epidemiological situation in humans in Europe, 2014/15. *J. Clin. Virol.* 82, 9–16. doi: 10.1016/j.jcv.2016.06.010
- Allen, M. D., Buckle, A. M., Cordell, S. C., Lowe, J., and Bycroft, M. (2003). The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macro H2A. *J. Mol. Biol.* 330, 503–511. doi: 10.1016/s0022-2836(03)00473-x
- Anang, S., Subramani, C., Nair, V. P., Kaul, S., Kaushik, N., Sharma, C., et al. (2016). Identification of critical residues in hepatitis E virus macro domain involved in its interaction with viral methyltransferase and ORF3 proteins. *Sci. Rep.* 6:25133. doi: 10.1038/srep25133
- Ankavay, M., Montpelliier, C., Sayed, I. M., Saliou, J. M., Wychowski, C., Saas, L., et al. (2019). New insights into the ORF2 capsid protein, a key player of the hepatitis E virus lifecycle. *Sci. Rep.* 9:6243. doi: 10.1038/s41598-019-42737-2
- Aslan, A. T., and Balaban, H. Y. (2020). Hepatitis E virus: epidemiology, diagnosis, clinical manifestations, and treatment. *World J. Gastroenterol.* 26, 5543–5560. doi: 10.3748/wjg.v26.i37.5543
- Borkakoti, J., Ahmed, G., and Kar, P. (2016). Report of a novel C1483W mutation in the hepatitis E virus polymerase in patients with acute liver failure. *Infect. Genet. Evol.* 44, 51–54. doi: 10.1016/j.meegid.2016.06.030
- Borkakoti, J., Ahmed, G., Rai, A., and Kar, P. (2017). Report of novel H105R, D29N, V27A mutations in the methyltransferase region of the HEV genome in patients with acute liver failure. *J. Clin. Virol.* 91, 1–4. doi: 10.1016/j.jcv.2017.03.017
- Cao, D., Ni, Y. Y., and Meng, X. J. (2018). Substitution of amino acid residue V1213 in the helicase domain of the genotype 3 hepatitis E virus reduces virus replication. *Virology* 15:32. doi: 10.1186/s12985-018-0943-5
- Cordoba, L., Huang, Y. W., Opriessnig, T., Harral, K. K., Beach, N. M., Finkelstein, C. V., et al. (2011). Three amino acid mutations (F51L, T59A, and S390L) in the capsid protein of the hepatitis E virus collectively contribute to virus attenuation. *J. Virol.* 85, 5338–5349. doi: 10.1128/JVI.02278-10
- Debing, Y., Gisa, A., Dallmeier, K., Pischke, S., Bremer, B., Manns, M., et al. (2014). A mutation in the hepatitis E virus RNA polymerase promotes its replication and associates with ribavirin treatment failure in organ transplant recipients.

- Gastroenterology* 147, 1008–1011.e7; quiz e1015–1006. doi: 10.1053/j.gastro.2014.08.040
- Debing, Y., Ramiere, C., Dallmeier, K., Piorkowski, G., Traub, M. A., Lebosse, F., et al. (2016). Hepatitis E virus mutations associated with ribavirin treatment failure result in altered viral fitness and ribavirin sensitivity. *J. Hepatol.* 65, 499–508. doi: 10.1016/j.jhep.2016.05.002
- Devhare, P., Sharma, K., Mhaindarkar, V., Arankalle, V., and Lole, K. (2014). Analysis of helicase domain mutations in the hepatitis E virus derived from patients with fulminant hepatic failure: effects on enzymatic activities and virus replication. *Virus Res.* 184, 103–110. doi: 10.1016/j.virusres.2014.02.018
- Ding, Q., Heller, B., Capuccino, J. M., Song, B., Nimgaonkar, I., Hrebikova, G., et al. (2017). Hepatitis E virus ORF3 is a functional ion channel required for release of infectious particles. *Proc. Natl. Acad. Sci. U. S. A.* 114, 1147–1152. doi: 10.1073/pnas.1614955114
- Garcia, N., Hernandez, M., Gutierrez-Boada, M., Valero, A., Navarro, A., Munoz-Chimeno, M., et al. (2019). Occurrence of hepatitis E virus in pigs and pork cuts and organs at the time of slaughter, Spain, 2017. *Front. Microbiol.* 10:2990. doi: 10.3389/fmicb.2019.02990
- Gavilan, A. M., Diez-Fuertes, F., Sanz, J. C., Castellanos, A. M., Lopez-Perea, N., Jimenez, S. M., et al. (2022). Increase of diversity of mumps virus genotype G SH variants circulating among a highly immunized population: Spain, 2007–2019. *J. Infect. Dis.* 176. doi: 10.1093/infdis/jiac176
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* 17, 4713–4730. doi: 10.1093/nar/17.12.4713
- Graff, J., Nguyen, H., Yu, C., Elkins, W. R., St Claire, M., Purcell, R. H., et al. (2005). The open reading frame 3 gene of hepatitis E virus contains a cis-reactive element and encodes a protein required for infection of macaques. *J. Virol.* 79, 6680–6689. doi: 10.1128/JVI.79.11.6680-6689.2005
- Graff, J., Torian, U., Nguyen, H., and Emerson, S. U. (2006). A bicistronic subgenomic mRNA encodes both the ORF2 and ORF3 proteins of hepatitis E virus. *J. Virol.* 80, 5919–5926. doi: 10.1128/JVI.00046-06
- Graff, J., Zhou, Y. H., Torian, U., Nguyen, H., St Claire, M., Yu, C., et al. (2008). Mutations within potential glycosylation sites in the capsid protein of hepatitis E virus prevent the formation of infectious virus particles. *J. Virol.* 82, 1185–1194. doi: 10.1128/JVI.01219-07
- Gu, Y., Tang, X., Zhang, X., Song, C., Zheng, M., Wang, K., et al. (2015). Structural basis for the neutralization of hepatitis E virus by a cross-genotype antibody. *Cell Res.* 25, 604–620. doi: 10.1038/cr.2015.34
- Hall, M. C., and Matson, S. W. (1999). Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* 34, 867–877. doi: 10.1046/j.1365-2958.1999.01659.x
- Ikram, A., Hakim, M. S., Zhou, J. H., Wang, W., Peppelenbosch, M. P., and Pan, Q. (2018). Genotype-specific acquisition, evolution and adaptation of characteristic mutations in hepatitis E virus. *Virulence* 9, 121–132. doi: 10.1080/21505559.2017.1358349
- Inoue, J., Takahashi, M., Mizuo, H., Suzuki, K., Aikawa, T., Shimosegawa, T., et al. (2009). Nucleotide substitutions of hepatitis E virus genomes associated with fulminant hepatitis and disease severity. *Tohoku J. Exp. Med.* 218, 279–284. doi: 10.1620/tjem.218.279
- Kabrane-Lazizi, Y., Meng, X. J., Purcell, R. H., and Emerson, S. U. (1999). Evidence that the genomic RNA of hepatitis E virus is capped. *J. Virol.* 73, 8848–8850. doi: 10.1128/JVI.73.10.8848-8850.1999
- Khudyakov, Y. E., Favorov, M. O., Jue, D. L., Hine, T. K., and Fields, H. A. (1994). Immunodominant antigenic regions in a structural protein of the hepatitis E virus. *Virology* 198, 390–393. doi: 10.1006/viro.1994.1048
- Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* 72, 2197–2206. doi: 10.1099/0022-1317-72-9-2197
- Koonin, E. V., Gorbalenya, A. E., Purdy, M. A., Rozanov, M. N., Reyes, G. R., and 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. doi: 10.1073/pnas.89.17.8259
- Leung, A. K. L., Griffin, D. E., Bosch, J., and Fehr, A. R. (2022). The conserved macrodomain is a potential therapeutic target for coronaviruses and Alphaviruses. *Pathogens* 11: 94. doi: 10.3390/pathogens11010094
- Lhomme, S., Garrouste, C., Kamar, N., Saune, K., Abravanel, F., Mansuy, J. M., et al. (2014). Influence of polyproline region and macro domain genetic heterogeneity on HEV persistence in immunocompromised patients. *J. Infect. Dis.* 209, 300–303. doi: 10.1093/infdis/jit438
- Li, C., Debing, Y., Jankevicius, G., Neyts, J., Ahel, I., Coutard, B., et al. (2016). Viral macro domains reverse protein ADP-Ribosylation. *J. Virol.* 90, 8478–8486. doi: 10.1128/JVI.00705-16
- Magden, J., Takeda, N., Li, T., Auvinen, P., Ahola, T., Miyamura, T., et al. (2001). Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J. Virol.* 75, 6249–6255. doi: 10.1128/JVI.75.14.6249-6255.2001
- Malik, A., Kumar, D., Khan, A. A., Khan, A. A., Chaudhary, A. A., Husain, S. A., et al. (2018). Hepatitis B virus precore G1896A mutation in chronic liver disease patients with HBeAg negative serology from North India. *Saudi J. Biol. Sci.* 25, 1257–1262. doi: 10.1016/j.sjbs.2016.05.004
- Mishra, N., Walimbe, A. M., and Arankalle, V. A. (2013). Hepatitis E virus from India exhibits significant amino acid mutations in fulminant hepatic failure patients. *Virus Genes* 46, 47–53. doi: 10.1007/s11262-012-0833-7
- Munoz-Chimeno, M., Barturen, S., Garcia-Lugo, M. A., Morago, L., Rodriguez, A., Galan, J. C., et al. (2022). Hepatitis E virus genotype 3 microbiological surveillance by the Spanish reference laboratory: geographic distribution and phylogenetic analysis of subtypes from 2009 to 2019. *Euro Surveill.* 27: 2100542. doi: 10.2807/1560-7917.ES.2022.27.23.2100542
- Munoz-Chimeno, M., Cenalmor, A., Garcia-Lugo, M. A., Hernandez, M., Rodriguez-Lazaro, D., and Avellon, A. (2020). Proline-rich Hypervariable region of hepatitis E virus: arranging the disorder. *Microorganisms* 8: 1417. doi: 10.3390/microorganisms8091417
- Nagashima, S., Takahashi, M., Jirintai, Tanaka, T., Yamada, K., Nishizawa, T., et al. (2011). A PSAP motif in the ORF3 protein of hepatitis E virus is necessary for virion release from infected cells. *J. Gen. Virol.* 92, 269–278. doi: 10.1099/vir.0.025791-0
- Nair, V. P., Anang, S., Subramani, C., Madhvi, A., Bakshi, K., Srivastava, A., et al. (2016). Endoplasmic reticulum stress induced synthesis of a novel viral factor mediates efficient replication of genotype-1 hepatitis E virus. *PLoS Pathog.* 12:e1005521. doi: 10.1371/journal.ppat.1005521
- Nan, Y., Wu, C., Zhao, Q., Sun, Y., Zhang, Y. J., and Zhou, E. M. (2018). Vaccine development against zoonotic hepatitis E virus: open questions and remaining challenges. *Front. Microbiol.* 9:266. doi: 10.3389/fmicb.2018.00266
- Nan, Y., Yu, Y., Ma, Z., Khattar, S. K., Fredericksen, B., and Zhang, Y. J. (2014). Hepatitis E virus inhibits type I interferon induction by ORF1 products. *J. Virol.* 88, 11924–11932. doi: 10.1128/JVI.01935-14
- Neuvonen, M., and Ahola, T. (2009). Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *J. Mol. Biol.* 385, 212–225. doi: 10.1016/j.jmb.2008.10.045
- Oechslin, N., Da Silva, N., Szkolnicka, D., Cantrelle, F. X., Hanoulle, X., Moradpour, D., et al. (2022). Hepatitis E virus RNA-dependent RNA polymerase is involved in RNA replication and infectious particle production. *Hepatology* 75, 170–181. doi: 10.1002/hep.32100
- Parvez, M. K. (2013). Molecular characterization of hepatitis E virus ORF1 gene supports a papain-like cysteine protease (PCP)-domain activity. *Virus Res.* 178, 553–556. doi: 10.1016/j.virusres.2013.07.020
- Parvez, M. K. (2015). The hepatitis E virus ORF1 'X domain' residues form a putative macrodomain protein/Appr-1-pase catalytic-site, critical for viral RNA replication. *Gene* 566, 47–53. doi: 10.1016/j.gene.2015.04.026
- Parvez, M. K. (2017). Mutational analysis of hepatitis E virus ORF1 "Y-domain": effects on RNA replication and virion infectivity. *World J. Gastroenterol.* 23, 590–602. doi: 10.3748/wjg.v23.i4.590
- Parvez, M. K., and Khan, A. A. (2014). Molecular modeling and analysis of hepatitis E virus (HEV) papain-like cysteine protease. *Virus Res.* 179, 220–224. doi: 10.1016/j.virusres.2013.11.016
- Purdy, M. A., Harrison, T. J., Jameel, S., Meng, X. J., Okamoto, H., Van der Poel, W. H. M., et al. (2017). ICTV virus taxonomy profile: Hepeviridae. *J. Gen. Virol.* 98, 2645–2646. doi: 10.1099/jgv.0.000940
- Raji, Y. E., Toung, O. P., Taib, N. M., and Sekawi, Z. B. (2022). Hepatitis E virus: an emerging enigmatic and underestimated pathogen. *Saudi J. Biol. Sci.* 29, 499–512. doi: 10.1016/j.sjbs.2021.09.003
- Rehman, S., Kapur, N., Durgapal, H., and Panda, S. K. (2008). Subcellular localization of hepatitis E virus (HEV) replicase. *Virology* 370, 77–92. doi: 10.1016/j.viro.2007.07.036
- Robinson, R. A., Burgess, W. H., Emerson, S. U., Leibowitz, R. S., Sosnovtseva, S. A., Tsarev, S., et al. (1998). Structural characterization of recombinant hepatitis E virus ORF2 proteins in baculovirus-infected insect cells. *Protein Expr. Purif.* 12, 75–84. doi: 10.1006/prep.1997.0817
- Rozanov, M. N., Koonin, E. V., and Gorbalenya, A. E. (1992). Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J. Gen. Virol.* 73, 2129–2134. doi: 10.1099/0022-1317-73-8-2129

- Santoni, D., Ghosh, N., and Saha, I. (2022). An entropy-based study on mutational trajectory of SARS-CoV-2 in India. *Infect. Genet. Evol.* 97:105154. doi: 10.1016/j.meegid.2021.105154
- Saraswat, S., Chaudhary, M., and Sehgal, D. (2019). Hepatitis E virus cysteine protease has papain like properties validated by in silico modeling and cell-free inhibition assays. *Front. Cell. Infect. Microbiol.* 9:478. doi: 10.3389/fcimb.2019.00478
- Schemmerer, M., Wenzel, J. J., Stark, K., and Faber, M. (2022). Molecular epidemiology and genotype-specific disease severity of hepatitis E virus infections in Germany, 2010–2019. *Emerg. Microbes Infect.* 11, 1754–1763. doi: 10.1080/22221751.2022.2091479
- Shafat, Z., Ahmed, A., Parvez, M. K., and Parveen, S. (2021). Role of "dual-personality" fragments in HEV adaptation-analysis of Y-domain region. *J. Genet. Eng. Biotechnol.* 19:154. doi: 10.1186/s43141-021-00238-8
- Smith, D. B., Izopet, J., Nicot, F., Simmonds, P., Jameel, S., Meng, X. J., et al. (2020). Update: proposed reference sequences for subtypes of hepatitis E virus (species Orthohepevirus A). *J. Gen. Virol.* 101, 692–698. doi: 10.1099/jgv.0.001435
- Smith, D. B., Simmonds, P., Izopet, J., Oliveira-Filho, E. F., Ulrich, R. G., Johne, R., et al. (2016). Proposed reference sequences for hepatitis E virus subtypes. *J. Gen. Virol.* 97, 537–542. doi: 10.1099/jgv.0.000393
- Takahashi, K., Okamoto, H., Abe, N., Kawakami, M., Matsuda, H., Mochida, S., et al. (2009). Virulent strain of hepatitis E virus genotype 3, Japan. *Emerg. Infect. Dis.* 15, 704–709. doi: 10.3201/eid1505.081100
- Tam, A. W., Smith, M. M., Guerra, M. E., Huang, C. C., Bradley, D. W., Fry, K. E., et al. (1991). Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* 185, 120–131. doi: 10.1016/0042-6822(91)90760-9
- Todt, D., Gisa, A., Radonic, A., Nitsche, A., Behrendt, P., Suneetha, P. V., et al. (2016). In vivo evidence for ribavirin-induced mutagenesis of the hepatitis E virus genome. *Gut* 65, 1733–1743. doi: 10.1136/gutjnl-2015-311000
- Vikram, T., and Kumar, P. (2018). Analysis of hepatitis E virus (HEV) X domain structural model. *Bioinformation* 14, 398–403. doi: 10.6026/97320630014398
- von Felden, J., Alric, L., Pischke, S., Aitken, C., Schlabe, S., Spengler, U., et al. (2019). The burden of hepatitis E among patients with haematological malignancies: a retrospective European cohort study. *J. Hepatol.* 71, 465–472. doi: 10.1016/j.jhep.2019.04.022
- Yamada, K., Takahashi, M., Hoshino, Y., Takahashi, H., Ichiyama, K., Nagashima, S., et al. (2009). ORF3 protein of hepatitis E virus is essential for virion release from infected cells. *J. Gen. Virol.* 90, 1880–1891. doi: 10.1099/vir.0.010561-0
- Yin, X., Ying, D., Lhomme, S., Tang, Z., Walker, C. M., Xia, N., et al. (2018). Origin, antigenicity, and function of a secreted form of ORF2 in hepatitis E virus infection. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4773–4778. doi: 10.1073/pnas.1721345115
- Zhou, Y. H., Purcell, R. H., and Emerson, S. U. (2005). A truncated ORF2 protein contains the most immunogenic site on ORF2: antibody responses to non-vaccine sequences following challenge of vaccinated and non-vaccinated macaques with hepatitis E virus. *Vaccine* 23, 3157–3165. doi: 10.1016/j.vaccine.2004.12.020