Part 2. Biotechnology

UDC 619:616.98:578.891:578.2”21:636.4:599.731.11 DOI 10.36016/JVMBBS-2023-9-1-2-3

COMPARATIVE ANALYSIS OF THE OPEN READING FRAMES PROTEIN GENES OF GENOTYPE 4 HEPATITIS E VIRUS IN SWINE AND WILD BOAR

Lymanska O. Yu.

National Scientific Center ‘Institute of Experimental and Clinical Veterinary Medicine’, Kharkiv, Ukraine, e-mail: olgaliman@ukr.net

Summary. The goal of this study was to determine the molecular diversity of the open reading frames (ORFs) ORF1, ORF2, ORF3 protein genes from full-length genomes of genotype 4 hepatitis E virus (HEV) from pigs and wild boars at protein and gene level. Statistical technique Shannon entropy was used for mutational analysis of ORF1–ORF3 protein genes to identify amino acid substitutions in the HEV-4 sequences isolated from pigs and wild boars that were most susceptible to mutations. Gene selective pressure for genes was estimated using Tajima’s neutrality test. The ORF regions of 11 swine and 11 wild boar genotype 4 HEV isolates with complete genomes from the GenBank database were analyzed comparatively. The total number of polymorphic sites was determined. Nonsynonymous (amino acid changing) and synonymous (amino acid preserving) substitutions were identified in ORF1, ORF2, ORF3 in swine and wild boar HEV-4 isolates. No evidence of recombination was found for ORFs in 11 swine HEV-4 isolates, ORF2, ORF3 in 8 wild boar HEV-4 isolates. However, a recombination fragment with a length of 430 nucleotides was detected in the ORF1 gene of 3 wild boar HEV-4 isolates. Positive D Tajima factors were determined for ORF1, ORF2, ORF3 genes of swine HEV-4 and ORF1, ORF2 genes of wild boar HEV-4. While a negative value of D Tajima’s factor was determined for ORF3 gene of wild boar HEV-4. Molecular characteristics showing principal distinctions between the open-reading frames of swine and wild boar genotype 4 hepatitis E virus were obtained. Wild boar ORF1 is characterized by lower nucleotide diversity π value (0.144) and higher number of segregated sites S value (1,688) comparing with higher π value (0.159) and lower S value (1,602) of swine ORF1. Positive values of D Tajima’s factor for ORF1, ORF2 ORF3 genes of swine HEV-4 and ORF1, ORF2 genes of wild boar HEV-4 show on positive selection of these genes. Negative value of D Tajima’s factor for ORF3 gene of wild boar HEV-4 indicates onto purifying selection decreasing variability in ORF3 gene of wild boar HEV-4. The largest number of amino acid variation sites (19.2%) was found for wild boar HEV-4 ORF3 followed by swine HEV-4 ORF3 (15.7%) comparing with other swine and wild boars HEV-4 ORFs

Keywords: mutational analysis, entropy analysis, Tajima’s neutrality test, positive selection, purifying selection

Introduction. Hepatitis E virus (HEV) is the causative agent of acute hepatitis — a dangerous liver disease with a mortality rate of about 1%. It is a major public health problem worldwide (Iaconelli et al., 2020). In contrast to known hepatitis viruses various animals are reservoirs for HEV (Pavio, Meng and Renou, 2010; Boadella, 2015).

Genotype 3 and genotype 4 HEV occupy a special place among the eight known genotypes (Wang and Meng, 2021) because of their zoonotic potential. Swine populations and wild boars are the main natural HEV reservoirs with the possibility of cross-species transmission (Fredriksson-Ahomaa, 2019; Salines, Andraud and Rose, 2017). The codon adaptation index for HEV3-4 exceeds 0.5 indicating a high adaptive potential to the host organism (Sun et al., 2020; Bouquet, Cherel, and Pavio, 2012). HEV3–4 is thought to be transmitted by ingestion of food from infected animals (Nan et al., 2017; Grierson et al., 2019).

HEV is a member of the family Hepeviridae, genus *Orthohepevirus*, species *Orthohepevirus A*. The HEV genome is represented by a single-stranded RNA molecule with a length of 7.2 knt consisting of three open reading frames (ORF). ORF1 has the most length of them, encodes non-structural proteins involved in HEV replication, and contains several functional domains (Ahmad, Holla and Jameel, 2011). ORF2 encodes the capsid protein which is a main structural component of the virion. The ORF2 protein is the main immunogenic target of neutralizing antibodies and exists in two forms — $ORF2^{S}$ (secreted form) and $ORF2^{C}$ (capsid associated form). In this case, translation is initiated from two different codons located at a distance of 15 amino acids (Yin et al., 2018). ORF3 encodes a multifunctional small phosphoprotein involved in HEV replication and pathogenesis (Kenney and Meng, 2019).

HEV as RNA virus exists as a mixture of quasispecies, i. e. closely related variants (Lauring and Andino, 2010). In spite of the similar transmission and the ability to cause chronic hepatitis, HEV-3 and HEV-4 differ in clinical manifestation and pathogenesis. In particular, humans and animals infected by HEV-4 show

significantly higher levels of alanine aminotransferase which is a marker of liver damage. HEV-4 causes fulminant hepatitis and early cirrhosis more often comparing with HEV-3 (Takahashi and Okamoto, 2014; Ohnishi et al., 2006; Perumpail et al., 2015).

The mutations and recombinations are the main mechanisms genomic diversity of viruses that can change their biological (contagiousness, virulence, etc.) or phenotypic properties (Domingo and Holland, 1997). Mutation process along with natural selection is a key factor of the evolution. HEV strains are characterized by the significant level of the genomic diversity despite the only one serotype existence (Okamoto, 2007).

The most HEV encoding regions are under the purifying (or negative) selection directed against arising mutations and changing amino acid sequences (Smith D. et al., 2012). The regions of HEV genome with a high level of amino acid substitutions are localized at the ORF2 N-end and ORF3 C-end. These regions are under the positive selection directing onto mutations spreading and fixation (Chen et al., 2012).

A region of overlapping reading frames for HEV3–4 is under the positive selection (Brayne et al., 2017). Among four main HEV genotypes (HEV1–4) HEV3–4 are the most diversified that may explain broad host range (Lara, Purdy and Khudyakov, 2014). Genome variability of viruses depends on its copying accuracy degree. The transcription is a source of HEV high mutation rate and genomic diversity (Van et al., 2016). HEV mutation rate of clinical isolates was found to be 1.5 nucleotide substitution per site in a year (Takahashi and Okamoto, 2014).

The ratio of the related rate of nonsynonymous mutations (dn) tothe related rate of synonymous mutations (ds) d*n*/d*s* is used to assess the variability deviation of virus genomes from the model of neutral molecular evolution (Kimura, 1991) and to determine selection mode. A ratio d*n*/d*s* is used by many researches for estimation of selection direction, selection strength for protein coding sequences and useful to distinguish various processes of evolution (Aziz et al., 2022; Gutierrez, Escalera-Zamudio and Pybus, 2019; Dasmeh et al., 2014). The ratio of numbers of synonymous substitutions per synonymous site / nonsynonymous substitutions per nonsynonymous site is used as a marker of the negative (< 1) or positive (> 1) selection.

This ratio was used for the characteristics of HEV quasispecies diversity at the acute phase of hepatitis E in solid-organ transpant patients. M and P capsid domains of HEV quasispecies in patients who developed chronic infection was found to be under the negative selection (Lhomme et al., 2012).

Statistical tests can be based on the estimation of difference between the number of single substitutions and the total number of substitutions (Fu and Li, 1993) or average value of pairwise nucleotide differences between sequences (Fu, 1997). A widely used efficient Tajima's test takes into account the number of variable sites and the average differences for large data sets in determining selection type (Tajima, 1989; Mohamed et al., 2019; Niczyporuk et al., 2020; Jadhav et al., 2020). The Tajima's test is based on the estimate of genetic diversity θ (substitutions per site) of a sequence alignment. Negative value of Tajima's indicates purifying selection, the value greater than zero indicates the positive selection (Yang, and Bielawski, 2000).

In the biological systems the processes of metabolism, energy and information exchange can be accompanied by both an increase and a decrease in entropy. A Shannon entropy is used for structural analysis of the biological systems condition at the macromolecules level and for determination of the degree of genetic variability at each amino acid or nucleotide position (Shannon, 1997). A higher Shannon entropy value at a sequence position indicates more variability in that position. A Shannon entropy value of zero indicates an invariant column of nucleotides / amino acid residues for all variants. Shannon entropy values were used for investigation of the different regions variability of genotypes 1, 3 and 4 HEV. The entropy values in X domain, RNA-dependent RNA polymerase domain, ORF2, ORF3 were determined to be the highest in HEV-3 and HEV-4 comparing with HEV-1 (Muñoz-Chimeno et al., 2022).

In this paper molecular characterization of three open reading frames in swine and wild boar HEV-4 on the macromolecules level was performed by computational methodologies.

Materials and methods. Full-length genomes for isolated from pig and wild boar genotype 4 HEV strains were obtained by searching the NCBI Nucleotide Database using the taxonomic identifier (txid) 1678143, along with associated metadata on host, country, and date of sampling. The parameters of the strains considered for the present study are listed in Table 1.

Table 1 — Parameters of isolated from swine and wild boars HEV-4 strains analyzed in the present study

	Strain/Isolate	GenBank record	Country	Year
Swine HEV-4	HB-S3	KX531115	China	2014
	HN-JY40	KM253769	China	2015
	CHN-SD-sHEV	KF176351	China	2011
	KM01	KJ155502	China	2010
	hb-3	GU361892	China	2008
	CHN-XJ-SW33	GU119960	China	2009
	bjsw1	GU206559	China	2008
	IND-SW-00-01	AY723745	India	2006
	swGX32	EU366559	China	2007
	BeSW67HEV4-2008	OM388298	Belgium	2008
	SS19	JX855794	China	2011

Table 1 — continuation

	Strain/Isolate	GenBank record	Country	Year
Wild boar HEV-4	G4HEV121-12cc	LC657084	Japan	2008
	2003-TL01	LC646471	Japan	2003
	JTF-Yamagu11	AB698654	Japan	2017
	CN-HuN2	MZ544007	China	2020
	wbJGF_08-1	AB602440	Japan	2008
	CN-GS3	MZ544006	China	2020
	CN-XJ7	MZ544005	China	2019
	CN-IM14	MZ544004	China	2019
	CN-JL23	MZ544003	China	2018
	CN-JL14	MZ544002	China	2018
	CN-CQ3	MZ544001	China	2019

The ORF genes sequences for the present study were divided into two datasets. Dataset 1 contained ORF 1 — ORF3 genes sequences from 11 swine HEV-4 strains. Dataset 2 consisted of ORF1–ORF3 genes sequences isolated from 11 HEV-4 strains of wild boar. Alignments for all two datasets were carried out using Molecular Evolutionary Genetics Analysis (MEGA) software (version 6.06) (Tamura et al., 2013). BioEdit (version 7.2.5) (Hall, 2013) software was used for mutational analysis of ORF1–ORF3 protein genes to determine the amino acid substitutions in the HEV-4 sequences isolated from pigs and wild boars. Gene selective pressure for genes was estimated using the Tajima's neutrality test by MEGA 6. Translation of nucleotide sequences in amino acid ones with following codon analysis was performed by BioEdit.

Shannon entropy is a useful quantification of diversity at a single position. Entropy plots for ORF1–ORF3 protein sequences, representing the amount of amino acid (and hence nucleotide) variability through each column in aligned sequences, were calculated by BioEdit.

Aligned ORF1–ORF3 gene sequences were screened for recombination with recombination detection program RDP4 (version 4.101) (Martin et al., 2015), using five available methods (RDP (Martin and Rybicki, 2000), GENECONV (Padidam, Sawyer and Fauquet, 1999), BootScan (Martin et al., 2005), MaxChi (Smith J., 1992), SiScan (Gibbs, Armstrong and Gibbs, 2000)) with default settings.

Results. More than thirty years have passed since the isolation of the first animal HEV strain (in swine), which was reported in 1990 (Reyes et al., 1990; Kordyum, 2001).

Analysis of mutations in ORF1–ORF3 protein genes. ORF1, ORF2, ORF3 gene sequences were restricted from full-length genomes of pig and wild boar genotype 4 HEV isolates. Nonsynonymous (amino acid-changing) and synonymous (amino acid-preserving) substitutions were identified in the ORF1, ORF2, ORF3 proteins of swine and wild boar genotype 4 HEV isolates (Table 2).

Table 2 — The number of polymorphic sites in aligned ORF1–ORF3 proteins of swine and wild boar HEV-4 comparing with complete length (in paranthesis) and position of hypervariavle region (HVR)

	ORF1	ORF2	ORF3
Swine HEV-4	91 (1,708) HVR: 719-789, 1,515-1,708	43 (623) HVR: 624-661	18 (114)
Wild boar HEV-4	78 (1,708) HVR: 717-789	34 (660)	22 (114)

Nonsynonymous substitutions in the ORF3 proteins of swine and wild boar genotype 4 HEV are summarized in Table 3.

Table 3 — The specific codon positions along with nonsynonymous (amino acid-changing) substitutions in the ORF3 protein of swine HEV-4 (left) and wild boar HEV-4 (right). The sequences GU361892 and LC646471 were used as a reference genotype for swine HEV-4 ORF3 and wild boar HEV-4 ORF3, respectively. Unique substitutions for ORF-3 of swine and wild boar HEV-4 are highlighted

Codon position	GenBank record	Amino acid residues substitution
Swine HEV-4		
2	EU366959, GU119960, GU206559	E (A)
32	AY723745	A (T)
34	EU366959, GU119960	A (V)
35	EU366959, GU119960	A (T)
39	KM253769	H (P)
68	AY723745	Q (R)
71	KF176351	Q (R)
73	GU119960, EU366959	P (Q)
74	AY723745, JX855794, OM388298	P (Q)
82	EU366959, GU119960, KX531115	G (D)
83	EU366959, GU119960	R (N)
	GU206559, AY723745, JX855794, KJ155502, KM253769, OM388298	R (S)
84	GU119960, KJ155502, KM253769	Q (R)
86	EU366959, GU206559	A (V)
89	JX855794	A (V)
93	AY723745, EU366959, JX855794	V (A)
94	AY723745	T (I)
102	GU119960	P (L)
104	KF176351, KJ155502, KM253769	V (A

Table 3 — continuation

Codon position	GenBank record	Amino acid residues substitution
Wild boar HEV-4		
34	LC657084, MZ544003	A (V)
39	AB698654, MZ544004	A (T)
45	LC657084	V (A)
67	AB698654, MZ544001	L (S)
69	AB698654, MZ544001, MZ544003 MZ544004, MZ544005, MZ544006, MZ544007	P (L)
72	MZ544006	Q (R)
73	MZ544002	P (L)
74	MZ544001, MZ544002, MZ544004, MZ544005, MZ544006, MZ544007	Q (P)
82	MZ544001	G (D)
83	MZ544001, MZ544007	S (R)
	MZ544006	S (P)
84	MZ544003, MZ544005	Q (R)
85	MZ544005	P (S)
86	AB698654, MZ544001, MZ544002, MZ544004, MZ544005, MZ544006, MZ544007	V (A)
88	AB698654	S (L)
89	MZ544003, MZ544005	A (V)
91	AB698654, MZ544001, MZ544002, MZ544003, MZ544004, MZ544005, MZ544006, MZ544007	P (L)
93	AB698654, MZ544002, MZ544003, MZ544007	V (A)
94	AB698654	T (I)
95	AB698654, MZ544001, MZ544002, MZ544003, MZ544004, MZ544006	N (S)
	MZ544007, LC657084	N (K)
100	MZ544003	P (L)
101	AB698654	L (P)
104	MZ544001, MZ544002, MZ544004, MZ544005, MZ544006, MZ544007	V (A)

Sequences were obtained from GenBank records and annotated by year of sampling. Data for the nonsynonymous substitutions in the ORF1 and ORF2 of swine and wild boar HEV-4 isolates are not shown.

Five unique substitutions were identified for swine ORF3 and nine ones for wild boar ORF3 in several HEV-4 isolates. These substitutions result in changing ORFs surface properties. For example, Q (P) amino acid residue substitution leads in decreasing polarity of wild boar HEV-4 ORF3 protein for 6 HEV isolates (position 74, Table 3) because P amino acid residue is characterized smaller polarity comparing with that of Q amino acid residue. N (S) and N (K) amino acid residues substitution results in increasing polarity of wild boar HEV-4 ORF3 protein for 8 HEV isolates (position 95, Table 3) because K and S amino acid residues are more polar comparing with that of N amino acid residue.

No evidence of recombination was found in the alignment for ORF1 gene sequence of pig isolates, ORF2, ORF3 gene sequences of pig and wild boar genotype 4 HEV isolates. But recombination was determined for ORF1 gene of wild boar isolates. Recombination fragment with length of 430 nucleotides (nt) was identified for three isolates from set of 11 ORF1 gene HEV-4 isolates. Fig. 1 graphically illustrates statistical evidence of recombination events between potential recombinant AB698654 (Japan, 2017), potential major parent MZ544005 (China, 2019), potential minor parent MZ544006 (China, 2020).

Analysis of Shannon entropy in ORF1–ORF3 protein sequences. Entropy measures the variability within site and assigns high score to highly variable sites and a lower score to less variable sites. Shannon entropy is a measure of the lack of information content (how could predict the position for a new incoming sequence) at each position in the alignment. By other words, entropy is a measure of the lack of predictability for an alignment position. For example, if any nucleotide (A, T, G or C) from four ones can be at position X with a frequency of 0.25, then information content has been reduced to 0, and the entropy is at maximum variability.

And, contrary, if there are N sequences in an alignment and at position K there is only one type nucleotide (for example, T) in all sequences, it can be assumed that there is a maximum information for position K. Assumption about nucleotide G at position K of another homologous sequence would be correct. That means a maximum information for position K, and the entropy in that case is 0.

A total of 92, 34, and 22 amino acid variation sites were identified by entropy analysis in dataset I for ORF1, ORF2 and ORF3 proteins of swine HEV-4. Variability is calculated as the entropy for each amino acid residue position. Entropy percentages for swine ORFs are as follow: ORF1 — 5.3% (91/1,708), ORF2 — 6.9% (43/623), and ORF3 — 15.7% (18/114).

A total of 78, 34, and 22 amino acid variation sites were identified by entropy analysis in datasets II for ORF1, ORF2 and ORF3 proteins of wild boar HEV-4. Entropy percentages for wild boar ORFs are as follow: ORF1 — 4.6% (78/1,708), ORF2 — 5.2% (34/660), and ORF3 — 19.2% (22/114). Entropy analysis revealed that wild boar ORF3 observed the largest variation (Fig. 2A) followed by swine ORF3 (Fig. 2B).

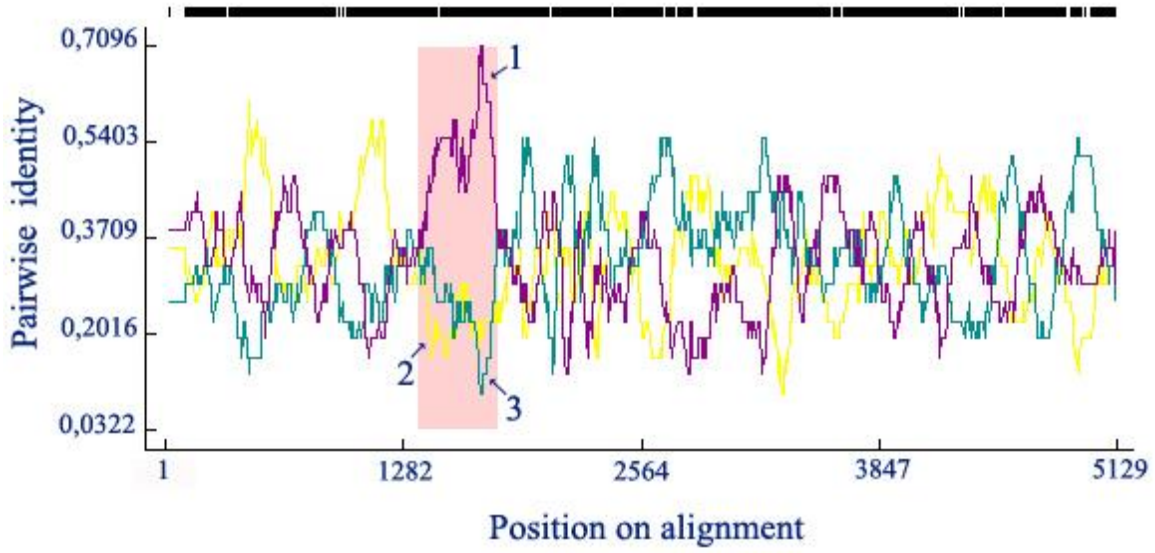

Figure 1. Plot of recombination events for ORF1gene of aligned three wild boar genotype 4 HEV isolates from set of 11 sequences. Different coloured lines indicate different sequence pairs: 1 — recombinant AB698654–minor parent MZ544006; 2 —major parent MZ544005–minor parent MZ544006; 3 — recombinant AB698654–major parent MZ544005. Pink rectangular area indicates ORF1 fragment with recombination length of 430 nucleotides of high identity between potential recombinant and closely related sequences.

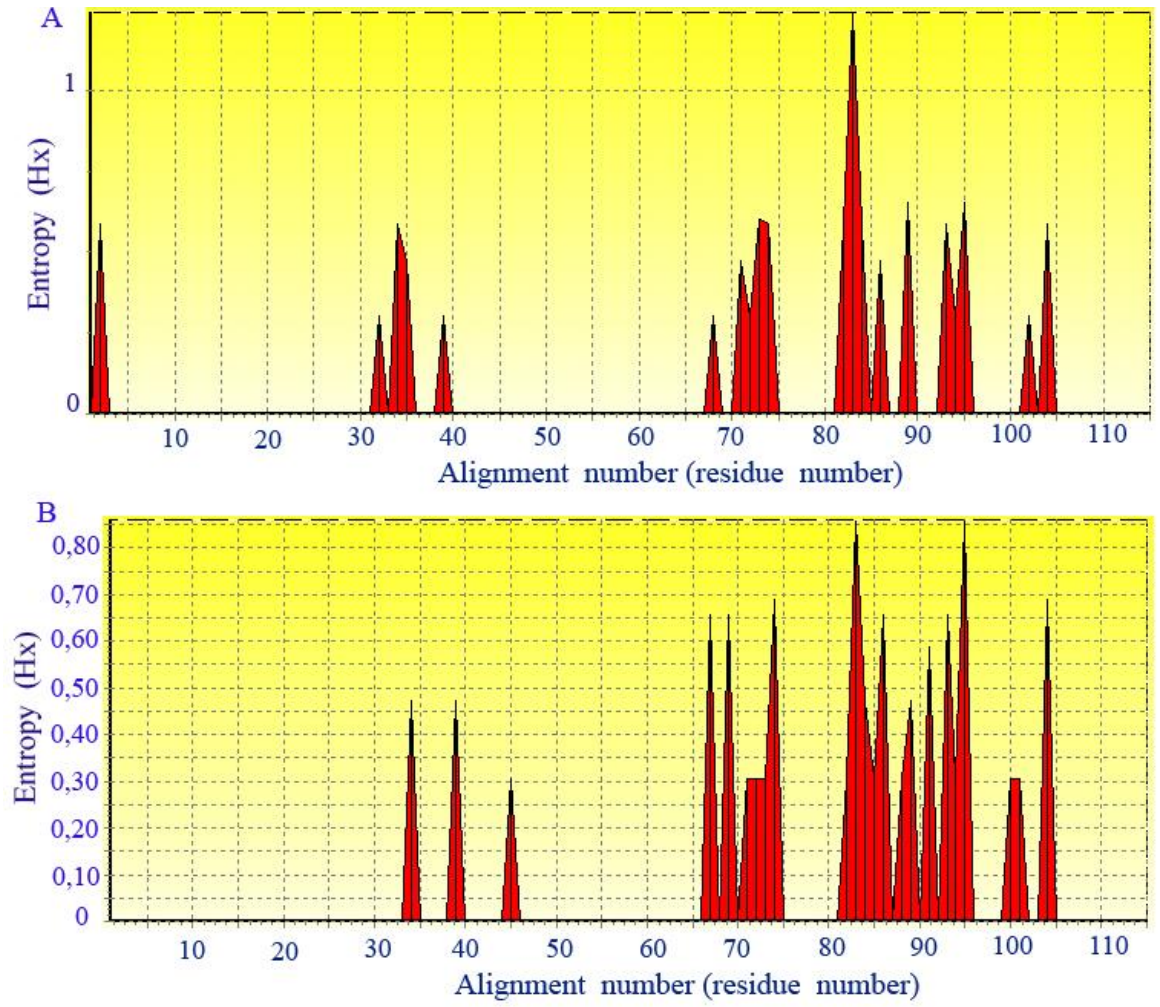

Figure 2. Entropy plots as a measure of diversity at each amino acid position for aligned amino acid sequences in ORF3 of swine HEV-4 (A) and wild boar HEV-4 (B).

Analysis of positive and purifying selection in ORF1–ORF3 protein genes. To identify ORF regions under positive or purifying selection, we estimated D value using Tajima's neutrality test. Parameter D indicates rating of correspondence nucleotide substitutions type to neutrality hypothesis for analyzed sequences. Deviation of observed diversity from model of neutral evolution may be determined by Tajima's test.

Veracious positive D factor values may indicate on sharp decreasing virus population or compensatory selection. ORF1, ORF2, ORF3 genes of swine HEV-4 and ORF1, ORF2 genes of wild boar HEV-4 consisted of genes under positive selection as indicated by positive D factors in Table 4, respectively. The selection pressure revealed the prevalence of positively selected sites in mentioned genes.

Table 4 — Parameters of mutational analysis of ORF1–ORF3 protein genes of HEV genotype 4 from pig and wild boar. Parameter D is the Tajima test statistic which indicates rating of correspondence nucleotide substitutions type to neutrality hypothesis for studied sequences

		m	S	p_s	Θ	π	D
Swine HEV-4	ORF1	11	1,602	0.361	0.123	0.159	1.409
	ORF2	11	558	0.281	0.096	0.119	1.163
	ORF3	11	43	0.124	0.042	0.045	0.340
Wild boar HEV-4	ORF1	11	1,688	0.330	0.112	0.144	1.360
	ORF2	11	512	0.258	0.088	0.109	1.193
	ORF3	11	48	0.139	0.048	0.046	–0.143
	ORF3	10	48	0.136	0.048	0.046	–0.278

Abbreviations: m — number of sequences, n — total number of sites, S — number of segregating sites (number of polymorphic sites in sequences), p_s — S/n, Θ — p_s/a_1, π — average nucleotide diversity.

However, ORF3 gene of wild boar HEV-4 was found to be under purifying selection as indicated by negative D factor, i. e. –0.143. Negative value of D Tajima's factor indicates onto recent population growth or purifying selection decreasing variability.

This suggests that the ORF1, ORF2 genes as of swine as of wild boar HEV-4 evolution is mainly driven by positive selection. Positive selection was obtained for ORF3 gene of swine HEV-4 too. While prevalence of purifying selection in ORF3 gene of wild boar HEV-4 was observed.

Estimated number of segregated sites (S) in swine and wild boar ORF2–ORF3 is in accordance with nucleotide diversity (π). The highest S value as for swine ORF2–ORF3 as wild boar ORF2–ORF3 is correlated with highest π value for swine and wild boar ORFs HEV, respectively. But quit different situation is for comparing S and π values in swine and wild boar ORF1. Wild boar ORF1 is characterized by lower π value (0.144) and higher S value (1,688) in comparison with higher π value (0.159) and lower S value (1,602) of swine ORF1.

Conclusions. New significant comparative information on the ORF1–ORF3 proteins of swine and wild boar genotype 4 HEV was obtained.

Positive values of D Tajima's factor for ORF1, ORF2, ORF3 genes of swine HEV-4 and ORF1, ORF2 genes of wild boar HEV-4 show on positive selection of these genes.

Negative value of D Tajima's factor for ORF3 gene of wild boar HEV-4 indicates onto purifying selection decreasing variability in ORF3 gene of wild boar HEV-4.

Genomic diversity in the ORFs protein genes of swine and wild boar HEV-4 isolates is quite different. For swine and wild boar ORF1 and ORF2 lower nucleotide diversity π corresponds to lower number of segregated sites S value. But wild boar ORF1 is characterized by lower nucleotide diversity π value (0.144) and higher number of segregated sites S value (1,688) comparing with higher π value (0.159) and lower S value (1,602) of swine ORF1, respectively.

Wild boar HEV-4 ORF3 observed the largest number of amino acid variation sites (19.2%) followed by swine HEV-4 ORF3 (15.7%) comparing with other swine and wild boars HEV-4 ORFs.

Further computational approaches are required to compare obtained results with swine and wild boar genotype 3 HEV.

References

Ahmad, I., Holla, R. P. and Jameel, S. (2011) 'Molecular virology of hepatitis E virus', *Virus Research*, 161(1), pp. 47–58. doi: 10.1016/j.virusres.2011.02.011.

Aziz, R., Sen, P., Beura, P. K., Das, S., Tula, D., Dash, M., Namsa, N. D., Deka, R. C., Feil, E. J., Satapathy, S. S. and Ray, S. K. (2022) 'Incorporation of transition to transversion ratio and nonsense mutations, improves the estimation of the number of synonymous and non-synonymous sites in codons', *DNA Research*, 29(4), p. dsac023. doi: 10.1093/dnares/dsac023.

Boadella, M. (2015) 'Hepatitis E in wild ungulates: A review', *Small Ruminant Research*, 128, pp. 64–71. doi: 10.1016/j.smallrumres.2015.03.007.

Bouquet, J., Cherel, P. and Pavio, N. (2012) 'Genetic characterization and codon usage bias of full-length hepatitis E virus sequences shed new lights on genotypic distribution, host restriction and genome evolution', *Infection, Genetics and Evolution*, 12(8), pp. 1842–1853. doi: 10.1016/j.meegid.2012.07.021.

Brayne, A. B., Dearlove, B. L., Lester, J. S., Kosakovsky Pond, S. L. and Frost, S. D. W. (2017) 'Genotype-specific evolution of hepatitis E virus', *Journal of Virology*, 91(9), pp. e02241–16. doi: 10.1128/JVI.02241-16.

Chen, X., Zhang, Q., He, C., Zhang, L., Li, J., Zhang, W., Cao, W., Lv, Y.-G., Liu, Z., Zhang, J.-X. and Shao, Z.-J. (2012)

'Recombination and natural selection in hepatitis E virus genotypes', *Journal of Medical Virology*, 84(9), pp. 1396–1407. doi: 10.1002/jmv.23237.

Dasmeh, P., Serohijos, A. W. R., Kepp, K. P. and Shakhnovich, E. I. (2014) 'The influence of selection for protein stability on dN/dS estimations', *Genome Biology and Evolution*, 6(10), pp. 2956–2967. doi: 10.1093/gbe/evu223.

Domingo, E. and Holland, J. J. (1997) 'RNA virus mutations and fitness for survival', *Annual Review of Microbiology*, 51(1), pp. 151–178. doi: 10.1146/annurev.micro.51.1.151.

Fredriksson-Ahomaa, M. (2019) 'Wild boar: A reservoir of foodborne zoonoses', *Foodborne Pathogens and Disease*, 16(3), pp. 153–165. doi: 10.1089/fpd.2018.2512.

Fu, Y.-X. (1997) 'Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection', *Genetics*, 147(2), pp. 915–925. doi: 10.1093/genetics/147.2.915.

Fu, Y. X. and Li, W. H. (1993) 'Statistical tests of neutrality of mutations', *Genetics*, 133(3), pp. 693–709. doi: 10.1093/genetics/133.3.693.

Gibbs, M. J., Armstrong, J. S. and Gibbs, A. J. (2000) 'Sister-Scanning: A Monte Carlo procedure for assessing signals in recombinant sequences', *Bioinformatics*, 16(7), pp. 573–582. doi: 10.1093/bioinformatics/16.7.573.

Grierson, S. S., McGowan, S., Cook, C., Steinbach, F. and Choudhury, B. (2019) 'Molecular and *in vitro* characterisation of hepatitis E virus from UK pigs', *Virology*, 527, pp. 116–121. doi: 10.1016/j.virol.2018.10.018.

Gutierrez, B., Escalera-Zamudio, M. and Pybus, O. G. (2019) 'Parallel molecular evolution and adaptation in viruses', *Current Opinion in Virology*, 34, pp. 90–96. doi: 10.1016/j.coviro.2018.12.006.

Hall, T. A. (2013) *BioEdit v 7.2. 3. Biological Sequence Alignment Editor for Win 95/98/NT/2K/XP7*. Carlsbad, California: Ibis Biosciences.

Iaconelli, M., Bonanno Ferraro, G., Mancini, P., Suffredini, E., Veneri, C., Ciccaglione, A. R., Bruni, R., Della Libera, S., Bignami, F., Brambilla, M., De Medici, D., Brandtner, D., Schembri, P., D'Amato, S. and La Rosa, G. (2020) 'Nine-year nationwide environmental surveillance of hepatitis E virus in urban wastewaters in Italy (2011–2019)', *International Journal of Environmental Research and Public Health*, 17(6), p. 2059. doi: 10.3390/ijerph17062059.

Jadhav, A., Zhao, L., Liu, W., Ding, C., Nair, V., Ramos-Onsins, S. E. and Ferretti, L. (2020) 'Genomic diversity and evolution of quasispecies in Newcastle disease virus infections', *Viruses*, 12(11), p. 1305. doi: 10.3390/v12111305.

Kenney, S. P. and Meng, X.-J. (2019) 'Hepatitis E virus genome structure and replication strategy', *Cold Spring Harbor Perspectives in Medicine*, 9(1), p. a031724. doi: 10.1101/cshperspect.a031724.

Kimura, M. (1991) 'The neutral theory of molecular evolution: A review of recent evidence.', *The Japanese Journal of Genetics*, 66(4), pp. 367–386. doi: 10.1266/jjg.66.367.

Kordyum, V. A. (2001) 'Viruses evolution — an attempt of non-linear prognosis', *Biopolymers and Cell*, 17(6), pp. 467–486. doi: 10.7124/bc.0005D6.

Lara, J., Purdy, M. A. and Khudyakov, Y. E. (2014) 'Genetic host specificity of hepatitis E virus', *Infection, Genetics and Evolution*, 24, pp. 127–139. doi: 10.1016/j.meegid.2014.03.011.

Lauring, A. S. and Andino, R. (2010) 'Quasispecies theory and the behavior of RNA viruses', *PLoS Pathogens*, 6(7), p. e1001005. doi: 10.1371/journal.ppat.1001005.

Lhomme, S., Abravanel, F., Dubois, M., Sandres-Saune, K., Rostaing, L., Kamar, N. and Izopet, J. (2012) 'Hepatitis E virus quasispecies and the outcome of acute hepatitis E in solid-organ transplant patients', *Journal of Virology*, 86(18), pp. 10006–10014. doi: 10.1128/JVI.01003-12.

Martin, D. and Rybicki, E. (2000) 'RDP: detection of recombination amongst aligned sequences', *Bioinformatics*, 16(6), pp. 562–563. doi: 10.1093/bioinformatics/16.6.562.

Martin, D. P., Posada, D., Crandall, K. A. and Williamson, C. (2005) 'A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints', *AIDS Research and Human Retroviruses*, 21(1), pp. 98–102. doi: 10.1089/aid.2005.21.98.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015) 'RDP4: Detection and analysis of recombination patterns in virus genomes', *Virus Evolution*, 1(1), p. vev003. doi: 10.1093/ve/vev003.

Mohamed, N. S., Ali Albsheer, M. M., Abdelbagi, H., Siddig, E. E., Mohamed, M. A., Ahmed, A. E., Omer, R. A., Muneer, M. S., Ahmed, A., Osman, H. A., Ali, M. S., Eisa, I. M. and Elbasheir, M. M. (2019) 'Genetic polymorphism of the N-terminal region in circumsporozoite surface protein of *Plasmodium falciparum* field isolates from Sudan', *Malaria Journal*, 18(1), p. 333. doi: 10.1186/s12936-019-2970-0.

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

Nan, Y., Wu, C., Zhao, Q. and Zhou, E.-M. (2017) 'Zoonotic hepatitis E virus: An ignored risk for public health', *Frontiers in Microbiology*, 8, p. 2396. doi: 10.3389/fmicb.2017.02396.

Niczyporuk, J. S., Kozdrun, W., Czekaj, H., Piekarska, K. and Stys-Fijoł, N. (2020) 'Isolation and molecular characterization of Fowl adenovirus strains in Black grouse: First reported case in Poland', *PLoS One*, 15(9), p. e0234532. doi: 10.1371/journal.pone.0234532.

Ohnishi, S., Kang, J., Maekubo, H., Arakawa, T., Karino, Y., Toyota, J., Takahashi, K. and Mishiro, S. (2006) 'Comparison of clinical features of acute hepatitis caused by hepatitis E virus (HEV) genotypes 3 and 4 in Sapporo, Japan', *Hepatology Research*, 36(4), pp. 301–307. doi: 10.1016/j.hepres.2006.08.002.

Okamoto, H. (2007) 'Genetic variability and evolution of hepatitis E virus', *Virus Research*, 127(2), pp. 216–228. doi: 10.1016/j.virusres.2007.02.002.

Padidam, M., Sawyer, S. and Fauquet, C. M. (1999) 'Possible emergence of new geminiviruses by frequent recombination', *Virology*, 265(2), pp. 218–225. doi: 10.1006/viro.1999.0056.

Pavio, N., Meng, X.-J. and Renou, C. (2010) 'Zoonotic hepatitis E: animal reservoirs and emerging risks', *Veterinary Research*, 41(6), p. 46. doi: 10.1051/vetres/2010018.

Perumpail, R. B., Ahmed, A., Higgins, J. P., So, S. K., Cochran, J. L., Drobeniuc, J., Mixson-Hayden, T. R. and Teo, C.-G. (2015) 'Fatal accelerated cirrhosis after imported HEV genotype 4 infection', *Emerging Infectious Diseases*, 21(9), pp. 1679–1681. doi: 10.3201/eid2109.150300.

Reyes, G. R., Purdy, M. A., Kim, J., Luk, K.-C., Young, L. M., Fry, K. E. and Bradley, D. W. (1990) 'Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis', *Science*, 247(4948), pp. 1335–1339. doi: 10.1126/science.2107574.

Salines, M., Andraud, M. and Rose, N. (2017) 'From the epidemiology of hepatitis E virus (HEV) within the swine reservoir to public health risk mitigation strategies: A comprehensive review', *Veterinary Research*, 48(1), p. 31. doi: 10.1186/s13567-017-0436-3.

Shannon, C. E. (1997) 'The mathematical theory of communication. 1963', *M. D. Computing: Computers in Medical Practice*, 14(4), pp. 306–317. PMID: 9230594.

Smith, D. B., Vanek, J., Ramalingam, S., Johannessen, I., Templeton, K. and Simmonds, P. (2012) 'Evolution of the hepatitis E virus hypervariable region', *Journal of General Virology*, 93(11), pp. 2408–2418. doi: 10.1099/vir.0.045351-0.

Smith, J. (1992) 'Analyzing the mosaic structure of genes', *Journal of Molecular Evolution*, 34(2), pp. 126–129. doi: 10.1007/BF00182389.

Sun, J., Ren, C., Huang, Y., Chao, W. and Xie, F. (2020) 'The effects of synonymous codon usages on genotypic formation of open reading frames in hepatitis E virus', *Infection, Genetics and Evolution*, 85, p. 104450. doi: 10.1016/j.meegid.2020.104450.

Tajima, F. (1989) 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism', *Genetics*, 123(3), pp. 585–595. doi: 10.1093/genetics/123.3.585.

Takahashi, M. and Okamoto, H. (2014) 'Features of hepatitis E virus infection in humans and animals in Japan: HEV infection in Japan', *Hepatology Research*, 44(1), pp. 43–58. doi: 10.1111/hepr.12175.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', *Molecular Biology and Evolution*, 30(12), pp. 2725–2729. doi: 10.1093/molbev/mst197.

Van Tong, H., Hoan, N. X., Wang, B., Wedemeyer, H., Bock, C.-T. and Velavan, T. P. (2016) 'Hepatitis E virus mutations: Functional and clinical relevance', *EBioMedicine*, 11, pp. 31–42. doi: 10.1016/j.ebiom.2016.07.039.

Wang, B. and Meng, X.-J. (2021) 'Structural and molecular biology of hepatitis E virus', *Computational and Structural Biotechnology Journal*, 19, pp. 1907–1916. doi: 10.1016/j.csbj.2021.03.038.

Yang, Z. and Bielawski, J. P. (2000) 'Statistical methods for detecting molecular adaptation', *Trends in Ecology & Evolution*, 15(12), pp. 496–503. doi: 10.1016/S0169-5347(00)01994-7.

Yin, X., Ying, D., Lhomme, S., Tang, Z., Walker, C. M., Xia, N., Zheng, Z. and Feng, Z. (2018) 'Origin, antigenicity, and function of a secreted form of ORF2 in hepatitis E virus infection', *Proceedings of the National Academy of Sciences*, 115(18), pp. 4773–4778. doi: 10.1073/pnas.1721345115.