Infection, Genetics and Evolution 68 (2019) 23–29

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human

Jian-hua Zhou[a,b,c], Xue-rui Li[c], Xi Lan[c], Sheng-Yi Han[a,c], Yi-ning Wang[c], Yonghao Hu[a,*], Qiuwei Pan[b,*]

[a] College of Veterinary Medicine, Gansu Agricultural University, Lanzhou, 730070, Gansu, China
[b] Department of Gastroenterology and Hepatology, Erasmus MC-University Medical Center, Room Na-1011, s-Gravendijkwal 230, NL-3015 CE, Rotterdam, the Netherlands
[c] State Key Laboratory of Veterinary Etiological Biology, Chinese Academy of Agricultural Sciences, Lanzhou Veterinary Research Institute, Lanzhou, 730046, Gansu, China.

ARTICLE INFO

Keywords:
Hepatitis E virus
Transmission
Codon usage pattern
Host
Natural selection

ABSTRACT

Hepatitis E virus (HEV) is an important pathogen causing public health burden. Swine has been recognized as a main reservoir. Interestingly, genotype 1 HEV only infects human; whereas genotype 3 and 4 are zoonotic. However, there is a lack of in-depth understanding in respect to the transmission from swine to human. Codon usage patterns generally participate in viral survival and fitness towards its hosts. We have analyzed codon usage patterns of the three open reading frames (ORFs) for 243 full-length genomes of HEV genotypes 1, 3 and 4. The divergence of synonymous codon usage patterns is different in each ORF for genotypes 1, 3 and 4, but the genotype-specific codon usage bias in genotype 1 is stronger than those of genotypes 3 and 4. In respect to genotypes 3 and 4, compared with strains isolated from human, HEV isolated from swine shows appreciable variation in adaptation of codon usages to human or swine. These results may help to understand the transmission and host adaptation of HEV genotypes 3 and 4 from swine to human.

1. Introduction

Hepatitis E virus (HEV) infection imposes substantial threat to public health (Meng, 2016). HEV is a non-enveloped virion and has a single-strand positive-sense RNA sequence. In the viral genome, there are three open reading frames (ORFs), namely ORF1 coding for the non-structural polyprotein, ORF2 coding for viral capsid protein and ORF3 coding for a small phosphoprotein (Holla et al., 2013). It is classified into the family *Hepeviridae* within the genus *Orthohepevirus* which contains four species (*Orthohepevirus A*, *B*, *C* and *D*) (Doceul et al., 2016). Within the proposed species *Orthohepevirus A*, at least four genotypes (1–4) among the currently recognized eight genotypes have been identified to infect human (Smith et al., 2016). Genotypes 1 and 2 have been only isolated from human, while genotypes 3 and 4 have been isolated from many mammals including human, swine, wild boar, goat, cattle, deer and yak (Smith et al., 2014). By genetic identification of HEV strains isolated from different animal species, this virus is broadening its host range and achieves cross-species infection (Meng, 2016). Some HEV strains originated from swine are thought to be spillovers in transmission from swine to human. For example, a human (HE-JA04-1911) and two swine HEV isolates (swJ8-5 and sw-J12-4) in Japan may be originated from UK via the importation of pigs since 1900 (Inoue et al., 2006). In France, some human HEV strains (i.e. TLS25) are related to swine strains previously identified in Europe (Legrand-Abravanel et al., 2009). Phylogenetic analyses indicate that the swine isolate (G8) closely resembles HEV derived from Chinese patients with acute hepatitis (Wang et al., 2002). It is interesting that HEV strains isolated from swine play noticeable role in cross-species infection, but there is a lack of in-depth information about the transmission of HEV from swine to human (Van der Poel et al., 2018).

Sequence mutations play a key role in viral emergence, but synonymous mutations are good candidates for selectively neutral evolution, due to not affecting the primary structure of protein. Nonrandom synonymous codon usage patterns strongly reflect the action of natural selection on 'silent' changes in coding sequence (Novoa and Ribas De Pouplana, 2012). Mutation biases or selection involved in synonymous codon usage patterns have been suggested as important factors determining the evolutionary changes of viruses which enable them to enhance viral survival and viral fitness to their hosts (Butt et al., 2016; Nasrullah et al., 2015; Wong et al., 2010; Zhou et al., 2013c). However,

---

[*] Corresponding authors.
E-mail addresses: yhh0817@126.com (Y. Hu), q.pan@erasmusmc.nl (Q. Pan).

https://doi.org/10.1016/j.meegid.2018.11.024
Received 15 July 2018; Received in revised form 22 November 2018; Accepted 30 November 2018

the viral and host factors that are involved in cross-species and zoonotic HEV infection are largely unclear.

In this study, based on a large set of publicly available full-length HEV genomes derived from human and swine, we analyzed genetic features related to synonymous codon usage patterns in different ORFs to understand new insights into HEV transmission from swine to human.

2. Materials and methods

2.1. Full-length genome data of HEV

The full-length genomes and coding sequence annotations of 243 HEV strains, which belong to genotypes 1, 3 and 4, were obtained from the National Center for Biotechnology (NCBI) GenBank database, accessed on 1st March 2018. The demographics of the selected strains are listed in Supplementary Table S1, including genotype 1 ($n = 31$), genotype 3 derived from human ($n = 80$), genotype 4 derived from human ($n = 50$), genotype 3 isolated from swine ($n = 56$), and genotype 4 isolated from swine ($n = 26$).

2.2. Parity rule 2 analysis

The Parity rule 2 (PR2) plot analysis was carried out for identifying the effects of mutation pressure and natural selection on codon usage of individual ORF of each HEV strain. PR2 plot is composed of the abscissa represented by AU bias [$A_3/(A_3 + U_3)$] and the ordinate represented by GC bias [$G_3/(G_3 + C_3)$] (Sueoka, 1999). A_3, U_3, G_3 and C_3 mean nucleotide A, U, G and C content at the third codon position of the four-codon amino acids (alanine, arginine, glycine, leucine, proline, serine, threonine and valine) in ORF, respectively. When both of AU bias and GC bias are equal to 0.5, it means where A = U and G = C (PR2), with no bias between the influence of the mutation and selection rates (substitution rates) (Sueoka, 1995, 1999).

2.3. The general codon usage patterns in the three ORFs for different genotypes

For avoiding influence of amino acid composition or the coding sequence lengths in calculating synonymous codon usage bias of different gene samples, the relative synonymous codon usage value (RSCU) (Sharp and Li, 1986) was calculated for the three ORFs of each HEV strain. Based on RSCU data for the specific ORF of different genotypes, principal component analysis (PCA) can compress the high-dimensional information about multiple HEV strains within the specific genotype into a two-dimensional map. This provides visualization of the differences in the general codon usage pattern for the specific ORF between different genotypes. To better identify synonymous codon usage divergence between different genotypes, the codon usage patterns of HEV ORFs 1–3 of genotype 1 were regarded as references, respectively. The usage divergence of each synonymous codon in each ORF between genotype 1 and other genotypes was quantified by the simple formula ($c = \sqrt{a^2 + b^2}$). In this formula, a_n is represented by the difference between $f_{1'}$ value of the specific synonymous codon in ORF of genotype 1 and $f_{1'}$ value of the corresponding synonymous codon in the same ORF of the given genotype (genotype 3 or genotype 4). b_n is represented by the difference between $f_{2'}$ value of this synonymous codon of genotype 1 and $f_{2'}$ value of this synonymous codon of the given genotype. The c value represents codon usage divergence of the specific synonymous codon in ORF between genotype1 and the given genotype.

2.4. Analysis of the overall codon usage bias for HEV ORFs

To better investigate the overall codon usage bias for HEV ORF without the confounding influences of amino acid composition and gene size, effective number of codons (ENC) analysis was performed. The range of ENC values is from 20 (extreme codon usage bias caused by selecting only one of the possible synonymous codons for the specific amino acid) to 61 (usage of the possible synonymous codons for the specific amino acid in random and equivalence), according to a previous study (Wright, 1990).

2.5. The similarity extent of codon usage between HEV ORF and host

Identifying the effect of the general codon usage pattern of host on shaping general codon usage pattern of HEV ORF was carried out by similarity index. The similarity index was calculated by *D (A,B)* formula (Zhou et al., 2013b).

$$R(A,B) = \frac{\sum_{i=1}^{59} a_i \times b_i}{\sqrt{\sum_{i=1}^{59} {a_i}^2 \times \sum_{i=1}^{59} {b_i}^2}}$$

$$D(A,B) = \frac{1 - R(A,B)}{2}$$

where *R(A,B)* is defined as a cosine value of an included angle between *A* and *B* special vectors representing the degree of similarity between HEV ORF and cattle at the aspect of the overall codon usage pattern, a_i is defined as the RSCU value for a specific codon in 59 synonymous codons in the specific ORF of different genotypes, b_i is termed as the RSCU value for the same codon of hosts (human or swine) (Zhou et al., 2013a; Zhou et al., 2013b). Of note, the 59 RSCU were calculated for human and swine from a codon usage database (Nakamura et al., 2000). *D(A,B)* represents the potential effect of the overall codon usage of different hosts on that of HEV ORFs. When *D(A,B)* value is closer to zero, the two codon usage patterns should share higher similarity extent of codon usage. When *D(A,B)* value is closer to 1.0, the two codon usage patterns represent the significant divergence.

3. Results

3.1. Biased nucleotide usages in four-fold degenerate codon family in HEV ORFs

To identify whether the biased codon selections function at ORFs 1–3 of HEV isolated from human and swine, the relationship between U and A composition and C and G composition in four-fold degenerate codon families were analyzed by PR2 plot (Fig. 1). A and U were used more frequently than G and C in four-fold degenerate codon families of HEV ORF1 (Fig. 1A). The preference towards codon selections for ORF1 of genotype 1 was generally stronger than those of genotypes 3 and 4 derived from either human or swine by evolutional selections, including natural selection. C and G were used more frequently than A and U in ORF2, and preference towards codon selections in genotype 1 was strongly shaped by evolutionary dynamics including mutation pressure and natural selection (Fig. 1B). Compared with nucleotide usage trends in ORF3 of genotypes 3 and 4 derived from human or swine, it is more stable in ORF3 of genotype 1 (Fig. 1C).

3.2. Genotype-specific codon usages in HEV ORFs

PCA can be used to compress the high-dimensional vectors into a two-dimensional map. Because the amount of the first axis (f_1') and the second axis (f_2') of ORF in each genotype accounted for > 85% of the data inertia (Table S2), PCA can provide a more convenient way to visualize differences in codon usage patterns in HEV ORFs 1–3. As shown in Fig. S1, the divergence of 59 synonymous codon usages in the three ORFs was displayed in different genotypes, respectively. Further scoring the extent of codon usage bias for the specific synonymous

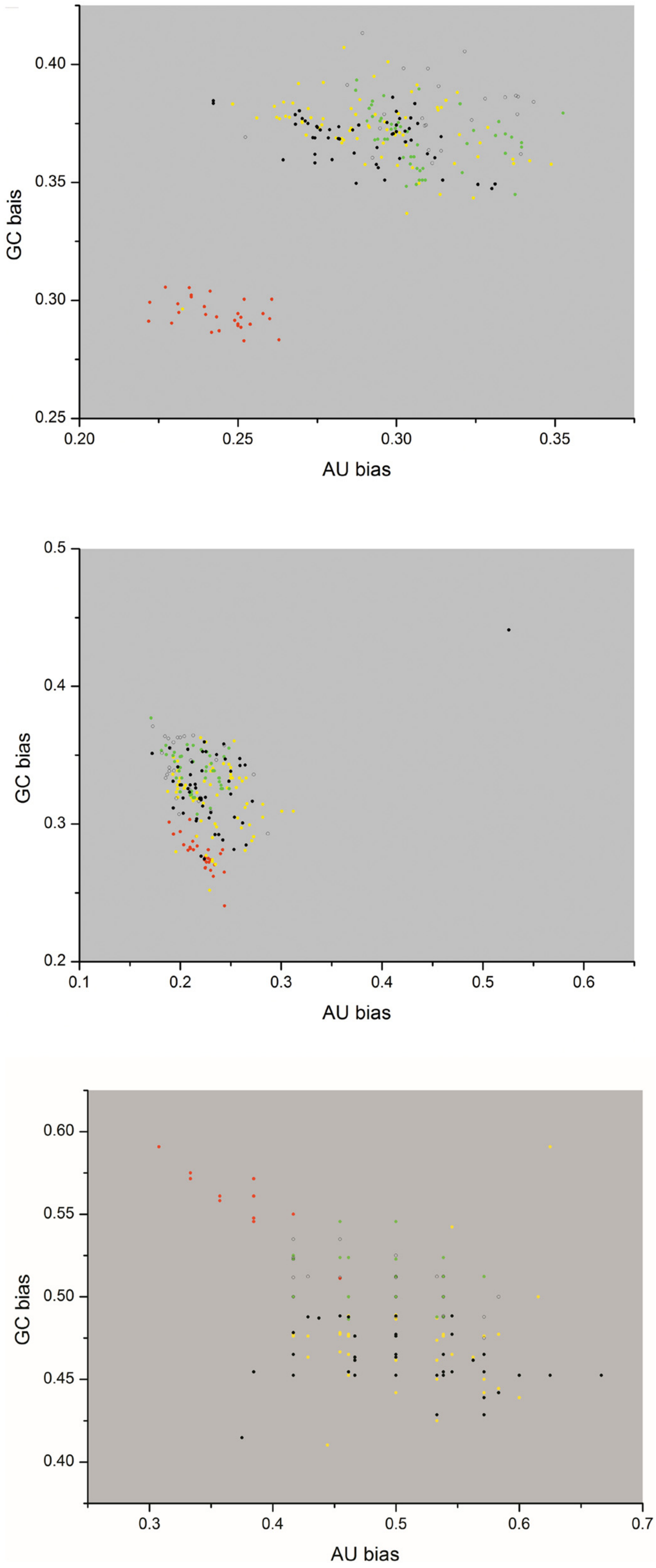


Fig. 1. Parity Rule 2 (PR2) bias plot for HEV. (A) PR2 biasness plot is calculated for each HEV ORF1 of different genotypes. (B) PR2 biasness plot is calculated for each HEV ORF2 of different genotypes. (C) PR2 biasness plot is calculated for each HEV ORF3 of different genotypes. Red dots: genotype 1 HEV strains; yellow dots: genotype 3 human HEV strains; green dots: genotype 4 human HEV strains; black dots: genotype 3 swine HEV strains; empty circles: genotype 4 swine HEV strains. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
The synonymous codons with obviously divergent usage pattern between genotype 1 and other genotypes.

	gt1 v.s. hgt 3	gt 1 v.s. hgt4	gt 1 v.s. sgt3	gt 1 v.s. sgt4
ORF1	UUG, CUA (Leu), GUC (Val), UCA, AGC (Ser), CCA (Pro), CGU, CGC, CGG, AGA (Arg)	GUU (Val), UCA (Ser), ACU, ACC (Thr), AAU, AAC (Asn), UGU, UGC (Cys), CGU, CGC (Arg)	UUG, CUA (Leu), GUU, GUC (Val), UCA (Ser), CCA (Pro), CGU, CGC, CGG (Arg)	UUG, CUA (Leu), AUC (Ile), GUC (Val), UCU, UCC (Ser), UAU, UAC (Tyr)
ORF2	AUC, AUA (Ile), ACU, ACC (Thr), CAU, CAC (His), AAA, AAG (Lys), UGU, UGC (Cys), GGU, GGC (Gly)	UUG, CUA (Leu), AUC (Ile), UCU, UCC (Ser), UGU, UGC (Cys)	UUU, UUC (Phe), AUC, AUA (Ile), ACU (Thr), AAA, AAG (Lys), UGU, UGC (Cys)	CAU, CAC (His), UGU, UGC (Cys), CGU, CGC (Arg)
ORF3	CUA (Leu), ACC, ACA, ACG (Thr), UAU (Tyr), CAA, CAG (Gln), GAG (Glu), CGC (Arg)	UCA (Ser), AAU, AAC (Asn), GAU, GAC (Asp), GAG (Glu)	ACC, ACA, ACG (Thr), UAU (Tyr), CAA, CAG (Gln), AAC (Asn), GAG (Glu), CGC (Arg)	GUC (Val), UCA (Ser), UAU (Tyr), GAU, GAC (Asp), GAG (Glu), CGU (Arg)

gt1 means genotype 1.
hgt3 means genotype 3 derived from human.
sgt3 means genotype 3 derived from swine.
hgt4 means genotype 4 derived from human.
sgt4 means genotype 4 derived from swine.

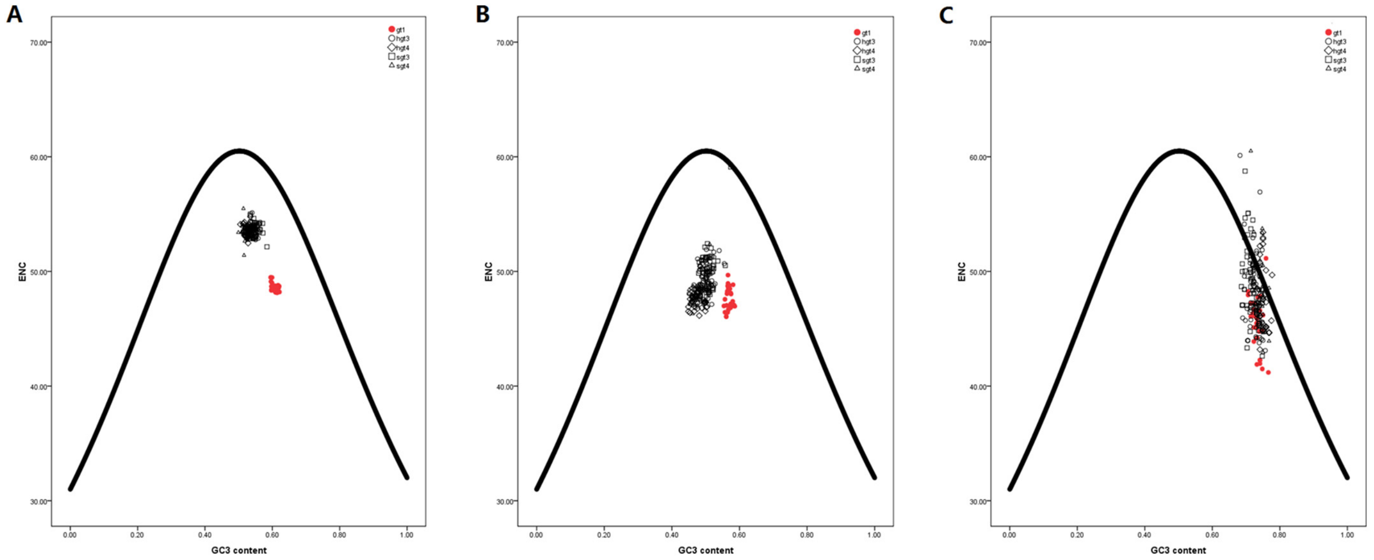


Fig. 2. The relationship between the overall codon usage patterns represented by ENC values and GC content at the third synonymous codon position (GC3 content) in the three ORFs of HEV. The solid curve is the expected line, which indicates that the overall codon usage pattern is only affected by mutation pressure caused by GC composition constraint. gt1: genotype 1; hgt3: genotype 3 derived from human; hgt4: genotype 4 means genotype 4 derived from human; sgt3: genotype 3 derived from swine; and sgt4: genotype 4 derived from swine.

codon in each ORF between genotype 1 and other genotypes was carried out based on the data of f_1' and f_2', and 59 synonymous codons in each ORF were analyzed via this identification, respectively (Table S3). According to the similarity extent of codon usage pattern of the three ORFs between genotype 1 and other genotypes (Table S2), the codons with obviously divergent usage pattern were found (Table 1). These codons were also regarded as genetic signs to reflect evolutional divergence between genotype 1 and other genotypes.

To analyze the role of mutation pressure, selection or drift caused by nucleotide composition constraint in the overall codon usage bias, we have assessed this by ENC versus GC3 contents. Most dots for ORFs 1 and 2 did not fall on the theoretical curve though tend to cluster to some degree, indicating that natural selection took part in the formation of the overall codon usage pattern. Of note, HEV ORFs 1 and 2 of genotype 1 were obviously separated from those of genotypes 3 and 4 (Fig. 2A and B). In contrast, the dots for ORF3 scattered to a large degree and strode across the theoretical curve, indicating that the formation of the overall codon usage pattern is mainly caused by mutation pressure due to nucleotide usage at the third position of codon in random (Fig. 2C).

3.3. Different codon usage adaptation of HEV ORFs to host

The similarity extent of codon usage between human and swine is 0.004, indicating that human and swine share a very high similarity extent of codon usage. The similarity extent of codon usage between HEV ORF1 and human is the highest (Fig. 3A), while it is the lowest between HEV ORF3 and human (Fig. 3C). Furthermore, compared with the codon usage adaptation of genotype 1 HEV ORF1 to human, the codon usage patterns of genotypes 3 and 4 HEV ORF1 have higher similarity with that of human (p value < .001) (Fig. 3A). Compared with the similarity extent of codon usage in ORFs 2 and 3 between genotype 1 HEV and human, the similarity extent between other genotypes and human highly varied (Fig. 3B and C). As for codon usage adaptation of HEV ORFs to swine, the codon usages of genotype 1 HEV ORFs represented the relatively stable models, but those of genotypes 3 and 4 showed the fluctuations with wide spectrum (Fig. 4). Compared with similarity extent of codon usage of HEV ORF 3 between genotype 1 HEV and swine (Fig. 4C), it is interesting that genotype 1 HEV ORFs 1 and 2 generally represent the better codon usage adaptation to swine than those of genotypes 3 and 4 (Fig. 4A and B). The codon usage adaptation of HEV ORFs to the two hosts indicates that codon usage patterns of genotype 1 HEV ORFs are strongly limited in either human or swine, while codon usage patterns of genotypes 3 and 4 HEV ORFs represent the relatively high dispersal. Interestingly, the similarity extent of codon usage between genotypes 3 and 4 HEV ORFs and human is stronger than that between genotypes 3 and 4 HEV ORFs and swine. These results seem paradoxical, but further indicate that genotypes 3 and 4 HEV strains derived from either human or swine have more diverse codon usage patterns in ORFs than those of genotype 1.

4. Discussion

In this study, we have performed a systematic approach to estimate the role of synonymous codon usage patterns in HEV ORFs in cross-species transmission. Obvious divergence of codon usage bias in four-fold degenerate codon family was identified in HEV ORFs between genotype 1 and genotypes 3 and 4. The genotype-specific codon usage, which results from the balance between mutational and natural selection, is an important evolutionary pathway at viral cross-species transmission. It is a type of synonymous codon usage bias which tends to be specific in viral coding sequences within a genotype and is shaped by multiple factors involved in itself and cellular environment. The clues of HEV transmission from swine to human have been indicated in respect to nucleotide identity and suggested the relationship between some specific nucleotide substitutions in HEV ORFs and cross-species transmission (Fu et al., 2011; Geng et al., 2011). It has also been proposed that mutagenesis of HEV ORFs is associated with the interactions between viral proteins and host cellulr proteins involved in metabolism and energy production (Ma et al., 2018; Ojha and Lole, 2016; Parvez, 2017).

HEV with high mutation rate is expanding host range and achieves cross-species infection, and this high mutation rate is also displayed by synonymous codon usage pattern of HEV genotypes 3 and 4 isolated from either human or swine. But it is noticed that long-term evolution of genotype 1 in human is strongly linked with synonymous codon usage patterns of HEV genotype 1. Genotype 1 HEV has a rather limited host range and is restricted to human, because attempts to experimentally infect other species including pigs, goats, and rats with genotype 1 HEV were not successful (Meng., 2016). This is an interesting

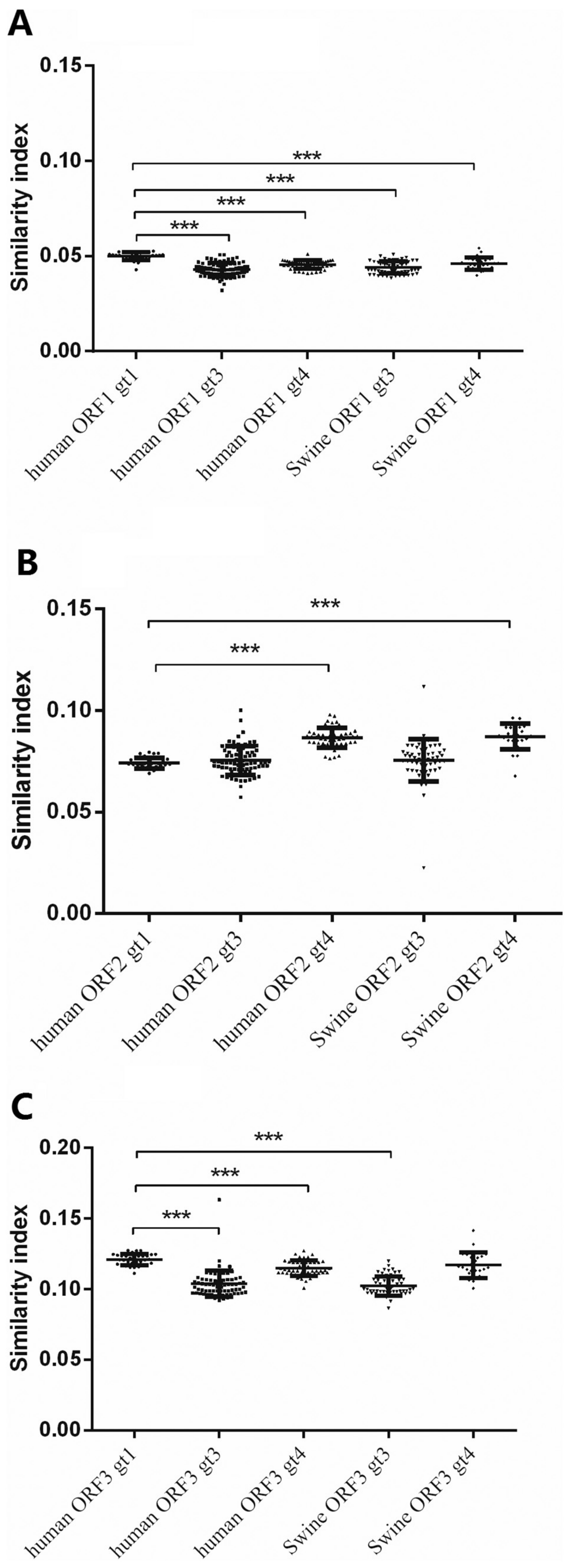
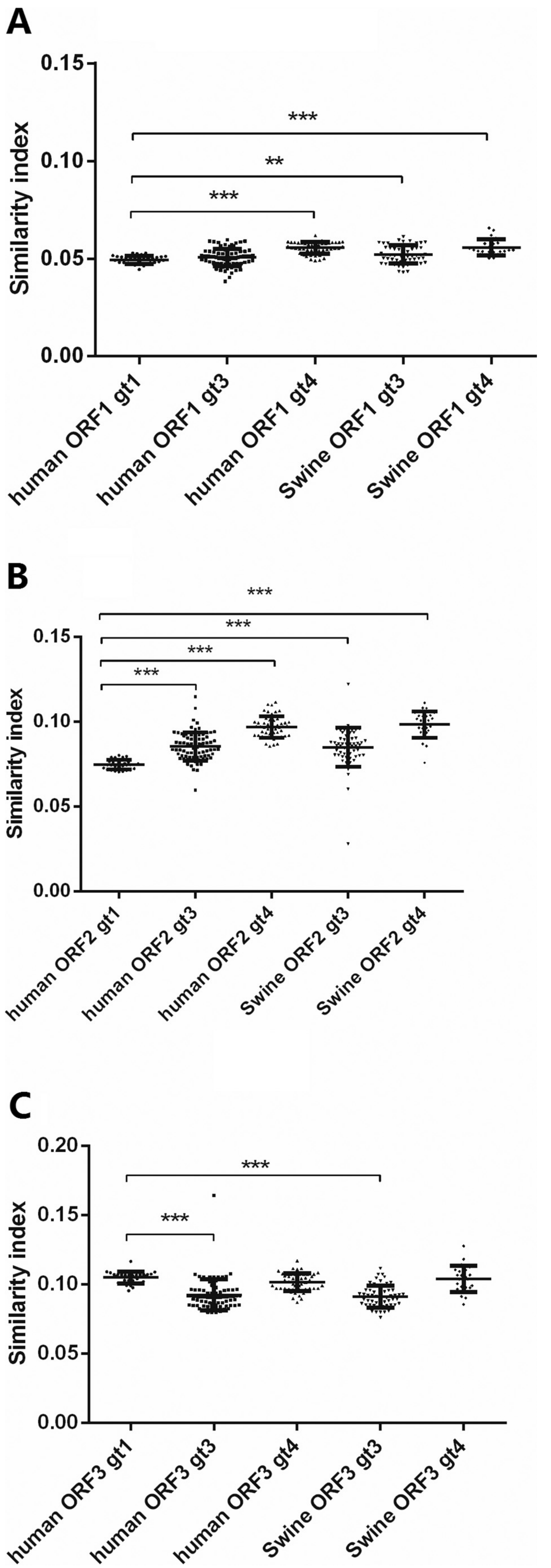


Fig. 3. Similarity index analysis of codon usage between HEV ORFs and human. (A) ORF1; (B) ORF2; (C) ORF3. The caption, e.g. 'human ORF1 gt1', means genotype 1 HEV ORF1 derived from human; ****p* value < 0.001.



(caption on next page)

Fig. 4. Similarity index analysis of codon usage between HEV ORFs and swine. (A) ORF1; (B) ORF2; (C) ORF3. The caption, e.g. 'human ORF1 gt1', means genotype 1 HEV ORF1 derived from human; **p value < .01; ***p value < 0.001.

aspect of HEV evolution, but there is a lack of information about the effects of mutation and natural selection on the codon usage of HEV ORFs and codon adaptation of viral ORFs to hosts. By analyzing nucleotide sequences in HEV genomes isolated from different animals, it has suggested that the *Orthopepevirus A* species is originated in human and subsequently may evolve into the zoonotic and human-restricted genotypes (Forni et al., 2018). The genotype-specific codon usage pattern further provides an evolutionary pathway in which HEV genotype 1 with neutral evolution (caused by synonymous mutations) is able to sustain its stable genetic characterization by strong natural selection. In contrast, the synonymous codon usage patterns of genotypes 3 and 4 HEV seem not to be explained by neutral evolution. The various synonymous codon usage patterns of genotypes 3 and 4 HEV reflect better codon usage adaptation to different hosts than those of genotype 1. The genotype-specific synonymous codon usage patterns of both genotype 1 and genotypes 3 and 4 can reflect the requirement of viral adaptation to host. It has been recognized that the synonymous codon usage in viral coding sequence functions as an important mediator in viral adaptation to host (Bahir et al., 2009). The restricted host infection and broad host infection of HEV serve as the strong natural selection for shaping the codon usage pattern of different genotypes HEV. Viral adaptation to host in codon usage is an evolutional feedback in the long-term evolution derived from host-range with HEV infection.

A previous in silico analysis has pointed out that geographical or host factors played weak roles in shaping the slight codon usage bias of HEV (Bouquet et al., 2012). To further assess the host factor related to natural selection in codon usages of HEV ORFs, the adaptation of codon usage in HEV ORFs to its hosts (human and swine) was estimated by $D(A,B)$ formula. This index is frequently used as a measure to assess the adaptation of viral genes to their hosts. We found a high similarity in codon usage pattern between human and swine. Host antiviral responses, e.g. interferon (IFN) response, essentially modulate viral infection. HEV infection elicits an active IFN-related antiviral response, initiated by the viral RNA sequence (Wang et al., 2018). Interestingly, each mammal owns a unique repertoire of IFN-stimulated genes and exert distinct responses, which potentially cause pathogen emergence and facilitate cross-species transmission (Shaw et al., 2017). Virus-host co-evolution has partially shaped the innate immune response, most frequently by placing antiviral genes under positive selection. The distinct host responses to different genotypes might drive the HEV evolutionary pathway. To evade immune system defense of host, the CpG and UpA dinucleotide frequencies play important roles in RNA virus replication and virulence. Synonymous codon usage pattern caused by dinucleotide usages represents selection pressures independently of coding capacity and profoundly influences host-pathogen interactions (Atkinson et al., 2014; Burns et al., 2009; Fros et al., 2017; Simon et al., 2017; Tulloch et al., 2014).

Besides host immune system, viral factors (especially viral genotypes and mutants) can modulate HEV life cycle and pathogenesis (van Tong et al., 2016). Compared with viral replication kinetics of genotypes 3 and 4, genotype 1 HEV has clear difference of either replication kinetics or antiviral response (Wu et al., 2018). Although HEV replication is controlled by the host immune system, viral factors (especially specific viral genomic variability and mutants) can modulate HEV replication, infection and pathogenesis(Okamoto, 2007; van Tong et al., 2016). The adaptation of codon usage of HEV ORFs to hosts is able to reflect a balance between mutation selection from virus and natural selection from hosts.

In summary, the genotype-specific codon usage bias in HEV genotype 1 is generally stronger than that of genotypes 3 and 4. Compared with the unique codon usage pattern of HEV genotype 1, the diverse synonymous codon usage patterns of genotypes 3 and 4 are evolutionary statues which may explain the transmission from swine to human. These results are helpful for better understanding the transmission and evolutionary adaptation of HEV to its hosts.

Disclosure of potential conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by National Natural Science foundation of China (No. 31302100).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.11.024.

References

Atkinson, N.J., Witteveldt, J., Evans, D.J., Simmonds, P., 2014. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic Acids Res. 42, 4527–4545.

Bahir, I., Fromer, M., Prat, Y., Linial, M., 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol. Syst. Biol. 5, 311.

Bouquet, J., Cherel, P., 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. Infect. Genet. Evol. 12, 1842–1853.

Burns, C.C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J., Kew, O., 2009. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. J. Virol. 83, 9957–9969.

Butt, A.M., Nasrullah, I., Qamar, R., Tong, Y., 2016. Evolution of codon usage in Zika virus genomes is host and vector specific. Emerg. Microbes Infect. 5, e107.

Doceul, V., Bagdassarian, E., Demange, A., Pavio, N., 2016. Zoonotic Hepatitis E Virus: Classi fi cation, Animal Reservoirs and Transmission Routes. Viruses 8.

Forni, D., Cagliani, R., Clerici, M., Sironi, M., 2018. Origin and dispersal of Hepatitis E virus. Emerg Microbes Infect 7, 11.

Fros, J.J., Dietrich, I., Alshaikhahmed, K., Passchier, T.C., Evans, D.J., Simmonds, P., 2017. CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. elife 6.

Fu, H., Wang, L., Zhu, Y., Geng, J., Li, L., Wang, X., Bu, Q., Zhuang, H., 2011. Analysing complete genome sequence of swine hepatitis E virus (HEV), strain CHN-XJ-SW13 isolated from Xinjiang, China: putative host range, and disease severity determinants in HEV. Infect. Genet. Evol. 11, 618–623.

Geng, J., Fu, H., Wang, L., Bu, Q., Liu, P., Wang, M., Sui, Y., Wang, X., Zhu, Y., Zhuang, H., 2011. Phylogenetic analysis of the full genome of rabbit hepatitis E virus (rbHEV) and molecular biologic study on the possibility of cross species transmission of rbHEV. Infect. Genet. Evol. 11, 2020–2025.

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

Inoue, J., Takahashi, M., Ito, K., Shimosegawa, T., Okamoto, H., 2006. Analysis of human and swine hepatitis E virus (HEV) isolates of genotype 3 in Japan that are only 81-83% similar to reported HEV isolates of the same genotype over the entire genome. J. Gen. Virol. 87, 2363–2369.

Legrand-Abravanel, F., Mansuy, J.M., Dubois, M., Kamar, N., Peron, J.M., Rostaing, L., Izopet, J., 2009. Hepatitis E virus genotype 3 diversity, France. Emerg. Infect. Dis. 15, 110–114.

Ma, X.X., Ma, Z., Pan, Q., 2018. The challenges of long-term transcriptional gene silencing by RNA viruses. Trends Biochem. Sci. 43, 649–650.

Meng, X.J., 2016. Expanding host range and cross-species infection of Hepatitis E virus. PLoS Pathog. 12, e1005695.

Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28, 292.

Nasrullah, I., Butt, A.M., Tahir, S., Idrees, M., Tong, Y., 2015. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. BMC Evol. Biol. 15, 174.

Novoa, E.M., Ribas De Pouplana, L., 2012. Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet. 28, 574–581.

Ojha, N.K., Lole, K.S., 2016. Hepatitis E virus ORF1 encoded non structural protein-host protein interaction network. Virus Res. 213, 195–204.

Okamoto, H., 2007. Genetic variability and evolution of hepatitis E virus. Virus Res. 127, 216–228.

Parvez, M.K., 2017. The hepatitis E virus nonstructural polyprotein. Future Microbiol 12, 915–924.

Sharp, P.M., Li, W.H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. 14, 7737–7749.

Shaw, A.E., Hughes, J., Gu, Q., Behdenna, A., Singer, J.B., Dennis, T., Orton, R.J., Varela, M., Gifford, R.J., Wilson, S.J., Palmarini, M., 2017. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. PLoS Biol. 15, e2004086.

Simon, D., Fajardo, A., Sonora, M., Delfraro, A., Musto, H., 2017. Host influence in the genomic composition of flaviviruses: a multivariate approach. Biochem. Biophys. Res. Commun. 492, 572–578.

Smith, D.B., Simmonds, P., Jameel, S., Emerson, S.U., Harrison, T.J., Meng, X.J., Okamoto, H., Van der Poel, W.H., Purdy, M.A., 2014. Consensus proposals for classification of the family Hepeviridae. J. Gen. Virol. 95, 2223–2232.

Smith, D.B., Simmonds, P., Izopet, J., Oliveira-Filho, E.F., Ulrich, R.G., Johne, R., Koenig, M., Jameel, S., Harrison, T.J., Meng, X.J., Okamoto, H., Van der Poel, W.H., Purdy, M.A., 2016. Proposed reference sequences for hepatitis E virus subtypes. J. Gen. Virol. 97, 537–542.

Sueoka, N., 1995. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. 40, 318–325.

Sueoka, N., 1999. Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G + C content of third codon position. Gene 238, 53–58.

Tulloch, F., Atkinson, N.J., Evans, D.J., Ryan, M.D., Simmonds, P., 2014. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. elife 3, e04531.

Van der Poel, W.H.M., Dalton, H.R., Johne, R., Pavio, N., Bouwknegt, M., Wu, T., Cook, N., Meng, X.J., 2018. Knowledge gaps and research priorities in the prevention and control of hepatitis E virus infection. Transbound. Emerg. Dis. 65 (Suppl. 1), 22–29.

van Tong, H., Hoan, N.X., Wang, B., Wedemeyer, H., Bock, C.T., Velavan, T.P., 2016. Hepatitis E Virus Mutations: Functional and Clinical Relevance. EBioMedicine 11, 31–42.

Wang, Y.C., Zhang, H.Y., Xia, N.S., Peng, G., Lan, H.Y., Zhuang, H., Zhu, Y.H., Li, S.W., Tian, K.G., Gu, W.J., Lin, J.X., Wu, X., Li, H.M., Harrison, T.J., 2002. Prevalence, isolation, and partial sequence analysis of hepatitis E virus from domestic animals in China. J. Med. Virol. 67, 516–521.

Wang, W., Wang, Y., Qu, C., Wang, S., Zhou, J., Cao, W., Xu, L., Ma, B., Hakim, M.S., Yin, Y., Li, T., Peppelenbosch, M.P., Zhao, J., Pan, Q., 2018. The RNA genome of hepatitis E virus robustly triggers an antiviral interferon response. Hepatology 67, 2096–2112.

Wong, E.H., Smith, D.K., Rabadan, R., Peiris, M., Poon, L.L., 2010. Codon usage bias and the evolution of influenza a viruses. Codon Usage Biases of Influenza Virus. BMC Evol. Biol. 10, 253.

Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87, 23–29.

Wu, X., Dao Thi, V.L., Liu, P., Takacs, C.N., Xiang, K., Andrus, L., Gouttenoire, J., Moradpour, D., Rice, C.M., 2018. Pan-genotype Hepatitis E virus replication in stem cell-derived hepatocellular systems. Gastroenterology 154, 663–674.

Zhou, J.H., Gao, Z.L., Zhang, J., Ding, Y.Z., Stipkovits, L., Szathmary, S., Pejsak, Z., Liu, Y.S., 2013a. The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. Infect. Genet. Evol. 14, 105–110.

Zhou, J.H., Zhang, J., Sun, D.J., Ma, Q., Chen, H.T., Ma, L.N., Ding, Y.Z., Liu, Y.S., 2013b. The distribution of synonymous codon choice in the translation initiation region of dengue virus. PLoS One 8, e77239.

Zhou, J.H., Zhang, J., Sun, D.J., Ma, Q., Ma, B., Pejsak, Z., Chen, H.T., Ma, L.N., Ding, Y.Z., Liu, Y.S., 2013c. Potential roles of synonymous codon usage and tRNA concentration in hosts on the two initiation regions of foot-and-mouth disease virus RNA. Virus Res. 176, 298–302.