DR. YIHAN LU (Orcid ID : 0000-0003-4651-9433)

Article type : Original Article

Distribution and phylogenetics of hepatitis E virus genotype 4 in humans and animals

Short title: Hepatitis E virus genotype 4

Bingzhe Li[1], Abram L. Wagner[2], Yujian Song[1], Xiangxiang Chen[1], Yihan Lu[1,3,4]

[1] Department of Epidemiology, School of Public Health, Fudan University, Shanghai 200032, China

[2] Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA

[3] Ministry of Education Key Laboratory of Public Health Safety (Fudan University), Shanghai 200032, China

[4] Global Health Institute, Fudan University, Shanghai 200032, China

Corresponding author: Dr. Yihan Lu, Fosun Tower, 131 Dong An Road, Shanghai 200032, China. E-mail: luyihan@fudan.edu.cn.

Alternate corresponding author： Bingzhe Li, Fosun Tower, 131 Dong An Road, Shanghai 200032, China. E-mail: bzli19@fudan.edu.cn

Author contribution

YL conceived and designed the study. BL, YS, and XC collected the sequence data and conducted the phylogenetic analysis. BL prepared the manuscript. YL and AW critically revised the manuscript. All authors read the manuscript and approved the final manuscript.

Key words

Hepatitis E virus; genotype 4; phylogenetic analysis; zoonosis

Impacts

- China and Japan remain endemic for HEV-4, whereas France has documented a high prevalence of HEV-4 in Europe.
- Globally, prevalent HEV-4 subtypes changed remarkably over the last 18 years.
- Increases in affected areas and animal hosts imply consistent cross-border and cross-species transmission.

Words: 3372

Abstract: 250

Tables: 4

Figures: 4

Supplementary tables: 3

Abstract

Background

Worldwide, hepatitis E virus (HEV) infection is considered a significant public health concern. In particular, HEV genotype 4 (HEV-4) has spread to more areas and host species. In this study, we describe the global distribution of HEV-4 and characterize HEV-4 subtypes by host, country, and year of isolation.

Methods

We retrospectively collected HEV-4 sequences available before December 31, 2019, in GenBank. HEV-4 and its subtypes were determined using phylogenetic comparison with HEV reference sequences. Information on the isolation of the sequences was extracted from the GenBank or original publications. Temporal, spatial, and host characteristics of the sequences were summarized and nucleotide similarity was calculated based on five amplified fragments within HEV genome, stratified by host, country, and year.

Results

A total of 2295 HEV-4 complete and partial nucleotide sequences were studied. The majority (92.7%) was isolated in China's mainland, Japan, Hong Kong, and France. A total of 20 animal hosts were documented, though swine remained predominant (71.7%). Globally, prevalent HEV-4 subtypes changed remarkably over the last 18 years. Subtype 4a, 4b, 4d, and 4h were most commonly isolated (80.3%). Subtypes 4c, 4e, 4f, 4g, and 4i remained limited in temporal distribution. High nucleotide similarities were observed between the sequences amplified in HEV ORF2, in the same and neighboring countries, and in similar animal hosts.

Conclusion

China and Japan are endemic for HEV-4, and have all the subtypes. In Europe, France has a high prevalence of HEV-4. Increases in affected areas and animal hosts imply consistent cross-border and cross-species transmission.

Key words

Hepatitis E virus; genotype 4; phylogenetic analysis; zoonosis

Introduction

Hepatitis E virus (HEV) is a food-borne or waterborne pathogen which can cause liver disease. HEV has caused numerous outbreaks of hepatitis in multiple low- and middle-income countries, and subsequently results in a high fatality rate among pregnant women in these geographical areas (Kmush, Wierzba, Krain, Nelson, & Labrique, 2013). The global burden of HEV was estimated to be approximately 20 million infections and 3.4 million symptomatic cases every year (Rein, Stevens, Theaker, Wittenborn, & Wiersma, 2012). However, the real burden may be underestimated, due to dynamic antibodies response, limited genotypes and sensitivity of assays ("EASL Clinical Practice Guidelines on hepatitis E virus infection," 2018). Currently, HEV is most prevalent in Eastern Asia and Southern Asia (WHO, 2019), where the seroprevalence of anti-HEV antibodies have varied from 10% to 50% and 10% to 40%, respectively (WHO, 2010).

Hepatitis E virus (HEV) is a non-enveloped, positive-sense single-stranded RNA virus which belongs to the Genus *Orthohepevirus* within the family of *Hepeviridae* (Purdy et al., 2017; D. B. Smith,

Simmonds, Members Of The International Committee On The Taxonomy Of Viruses Hepeviridae Study, et al., 2015). HEV genome is approximately 7.2 kb in length, which includes three open reading frames (ORFs) (Tam et al., 1991; Tsarev et al., 1992). ORF 1 encodes non-structural proteins consisting of several functional domains such as a methyltransferase domain (Met), Y domain (Y), papain-like cysteine protease (PCP), hypervariable region (HV), proline-rich domain (Pro), X-domain (X), helicase (Hel), and RNA-dependent RNA polymerase (RdRp), whereas ORF 2 encodes structural proteins (Graff, Torian, Nguyen, & Emerson, 2006; Yuchen Nan, Wu, Qin, & Zhou, 2017; Panda & Varma, 2013). ORF 3, which sometimes overlaps with ORF 1 and ORF 2, encodes a minor structural protein (Tam et al., 1991).

So far, HEV has been genetically classified in four species, Orthohepevirus A to D (Donald B. Smith & Peter Simmonds, 2018). In the species A, a total of eight genotypes have been identified, with hosts such as humans (HEV-1, 2, 3, 4, and 7), domestic pigs (HEV-3 and 4), rabbits (HEV-3), wild boars (*Sus scrofa*) (HEV-3, 4, 5, and 6), and camels (HEV-7 and 8) (Donald B. Smith & Peter Simmonds, 2018). Species B, C, and D have been identified in only non-human animals; however, two cases of hepatitis E patients infected with rat HEV (species C) were reported in Hong Kong in 2018 (Sridhar et al., 2018). HEV exhibits high genetic variability by rapid evolution due to its nature of RNA virus and selection pressure imposed by immune responses of zoonotic hosts (Sanjuán, Nebot, Chirico, Mansky, & Belshaw, 2010; van Tong et al., 2016), which could lead to increasing genetic diversity and more emerging genotypes or subtypes. HEV-3 and 4 are the most commonly studied zoonotic genotypes. HEV-3 is prevalent worldwide (Primadharsini, Nagashima, & Okamoto, 2019), whereas HEV-4 is mostly limited to Asia, including China, Japan, and countries in Southeastern Asia (Lu, Li, & Hagedorn, 2006). However, HEV-4 has been recently isolated in Europe (Hakze-van der Honing, van Coillie, Antonis, & van der Poel, 2011). It remains unexplained why HEV-3 and 4 have different geographical distributions. HEV-4 was first identified in China in 1999 (Wang et al., 1999) and has become the predominant genotype in China since 2000 (Zhang et al., 2011). In addition, it is noted that there is increasing numbers of indigenous cases in Europe (Asma, Stephanie, Eric, Camelia, & Anne-Marie, 2013; Colson et al., 2012; Fogeda, Avellón, Cilla, & Echevarría, 2009b; Garbuglia et al., 2013; Midgley, Vestergaard, Dalgaard, Enggaard, & Fischer, 2014; Ole et al., 2008; Tessé et al., 2012). Currently, a total of nine HEV-4 subtypes, 4a – 4i, have been determined (D. B. Smith et al.,

2020).

In addition, previous studies have provided a large number of nucleotide sequences that facilitate in-depth study of HEV. In this study, we retrospectively studied all HEV-4 sequences available in the GenBank to determine the global distribution of HEV-4 and characterized each subtype by host, country, and year of isolation.

Materials and methods

HEV-4 sequence retrieval and genotyping

In the GenBank (https://www.ncbi.nlm.nih.gov/nucleotide), we defined our search strategies as follows: ("Hepatitis E virus"[Organism] OR hepatitis e virus [All Fields]) OR ("Hepatitis E virus"[Organism] OR HEV [All Fields]). Nucleotide sequences with 1 nucleotide (nt) through 10000 nt in length that were released before December 31, 2019, were retrieved in the GenBank. Those sequences that were synthesized or patented were excluded from the study. By duplicate alignments with HEV reference strains (D. B. Smith, Simmonds, Jameel, et al., 2015), we determined if the retrieved sequences were HEV-4 by constructing a Neighbor-Joining tree and a maximum likelihood tree with bootstrap tests of 1000 replications using the Kimura-2-parameter method in MEGA 7.0 software (www.megasoftware.net). A total of 48 full-length reference sequences representing eight genotypes of species A, including 11 HEV-4 reference sequences, were utilized for comparison (Supplementary Table 1) (D. B. Smith et al., 2020). Subtypes of HEV-4 were determined by duplicate comparison with 48 HEV reference sequences and 11 HEV-4 reference sequences with the Neighbor-Joining method. If there was conflicting output, the maximum likelihood method was further utilized to confirm the subtype. HEV genotype and subtype were assigned by Smith *et al (D. B. Smith et al., 2020)*. We listed all the uncertain comparison outputs in Supplementary Table 2.

Information on the isolation of the retrieved sequences was directly extracted from GenBank, including species host, country, date of collection, isolate/strain name, ORF, and CDS product. If there was incomplete information in the GenBank, we further searched the original publications that had cited the sequences to complete the information. Regarding missing date of collection, we estimated the year of isolation as one year before the submitted date. If the date was displayed in a

time period, we used the middle year as date of collection. Duplicate sequences that shared identical nucleotide positions, isolate/strain names, hosts, countries and dates of collection were removed from the analysis. This search and selection procedure was detailed in Fig 1.

HEV-4 sequence datasets

In this study, we generated six datasets for phylogenetic analysis. In dataset 1, there was a total of 2295 HEV-4 strains with complete information about isolation. Of them, 83 strains were full-length sequences while the others were partial sequences with different nucleotides in length. Subsequently we described the global distribution of HEV-4 subtypes by host, country, and year of isolation. Each strain was determined by excluding duplicate sequences that shared identical isolate/strain name, host, country and date of collection.

In datasets 2-6, the 83 full-length sequences were segmented into the following five partial fragments; those partial sequences not belonging to the five partial fragments were excluded in the following analysis. Finally, we included HEV-4 sequences that matched the five partial fragments as follows: ORF 1 Met (279 nt, n=232, 171-449 nt in subtype 4a reference sequence: AB197673), ORF 1 RdRp (307 nt, n=446, 4292-4598 nt in AB197673), and ORF 2 (317 nt, n=754, 6026-6342 nt in AB197673; 150 nt, n=368, 6356-6505 nt in AB197673; and 537 nt, n=325, 6449-6985 nt in AB197673). Nucleotide similarities within and between groups stratified by host, country, and year of isolation were calculated using the Kimura-2-parameter method with bootstrap tests of 1000 replications. HEV-4 sequences that belonged to the same strain but were in different nucleotide positions within the HEV genome were kept separately in dataset 2-6.

Ethical Approval statement

This study included the HEV sequences available in the GenBank and involved no ethical issue. Thus, this study did not require ethical approval.

Results

Global distribution of HEV-4

HEV-4 was first isolated in a Chinese patient with acute hepatitis in 1992 (Wang et al., 1999) and has

been circulating for decades. A total of 2295 HEV-4 strains were identified in this study (Fig 1). Information on all sequences was listed in Supplementary Table 3, including year of isolation, host, country and region, subtype, and GenBank accession number. Temporally, isolation of HEV-4 strains were very limited before 2000. From 2000 through 2004, the number of isolated strains increased slightly but remained at a low level. Between 2005 and 2011, the number of isolated strains remained high every year, with 2007 being the peak. Then the number of strains showed a significant decline, and began to increase slightly over 100 in 2015 and remained steady until 2018 (Fig 2).

Geographically, HEV-4 strains have been isolated in Asia, Europe, and North America involving 22 countries and regions, suggesting extensive distribution across the world. The vast majority of strains have been isolated in Asia (96.8%, 2223/2295), including China's mainland (n=1745), Hong Kong (n=92), and Japan (n=243). In addition, sporadic cases have been identified in Europe, and North America, with the majority isolated in France (n=48) but only one each from the UK and the USA (Table 1). China and France had the largest number of isolated HEV-4 strains in Asia and Europe, respectively. Asian countries that have isolated HEV-4, such as South Korea, India, Viet Nam, Mongolia, and Indonesia, are all neighboring to China. In European countries, such as Denmark, Italy, Belgium, Spain, and Germany, which are geographically close to France have also reported HEV-4 infection (Fig 3). The findings suggested a geographically central distribution of HEV-4 in China and France, and circular distribution in the neighboring countries and regions.

So far, a total of 20 species have been reported to be the host of HEV-4, including humans (n=1274), swine (n=732), and other animals (n=289). Generally, the number of human strains was larger than swine strains in most subtypes; however, swine strains surpassed human ones in subtype 4d (n=210 vs. n=206) and 4e (n=24 vs. n=16). Of the total animal HEV-4 strains, 84.9% (867/1021) were isolated in China's mainland, covering all the subtypes of animal hosts. In Europe, swine and cynomolgus monkey (*Macaca fascicularis*) were the only animal hosts. In addition to swine, wild boar (*Sus scrofa*), deer (*Cervidae*), and other mammals, animal hosts of HEV-4 included birds such as crowned crane (*Grus japonensis*) (GenBank accession no., EF417589) and silver pheasant (*Lophura nycthemera*) (EF417590), and shellfish such as blood clam (*Scapharca subcrenata*) (KJ816336-KJ8163340, KJ816346, KJ816347) and Philippine clam (*Ruditapes philippinarum*) (KJ816335,

KJ816342, KJ816343).

A total of ten HEV-4 subtypes and clusters were described in this study (Table 1). Subtype 4a (n=642) was annually isolated with human, swine, and wild boar hosts; a strain was recently isolated in cattle in South Korea (MK795980). Moreover, 4a was frequently isolated in Asia, except a US strain (JX426082) in 2010. Similarly, subtype 4c (n=99), 4e (n=40), and 4g (n=48) appeared mainly in Asia (except a 4e British strain GQ913689 in 2009 and two 4g Swiss strains MK343062 and MK343063 in 2012 and 2015, respectively) and had human, swine, and wild boar hosts. However, 4c, 4e, and 4g have not been isolated after 2016, 2009, and 2015, respectively. Subtype 4b (n=396) and 4d (n=473) had a wide range of hosts, including rhesus monkey (*Macaca mulatta*), cynomolgus monkey (*Macaca fascicularis*), cattle (*Bos taurus*), sheep (*Ovis aries*), donkey, wild boar, tiger (Bengal tiger & Siberian tiger), and shellfish, in addition to human and swine, which were isolated continually in Europe and Asia. Similarly, subtype 4h (n=331) had a variety of hosts, such as cattle, sheep, goat (*Capra hircus*), donkey (*Equus asinus*), wild boar, rhesus monkey, tree shrew (*Tupaia belangeri*) and wild yak (*Bos grunniens*). Subtype 4i (n=190) had mammalian hosts (deer, bear (*Ursus thibetanus*), and leopard (*Panthera pardus*)) and birds; however, it was only isolated in Asia before 2017. Subtype 4f had only 6 strains in humans, and 3 strains in wild boars, in which the latest was isolated in Hong Kong in 2014 (KX752741). Subtype 4 (n=12) was distinct from other subtypes but remained unassigned (Donald B. Smith, Simmonds, Izopet, Oliveira-Filho, & Purdy, 2016) with human and swine hosts across Asia and Europe, of which the latest two strains were isolated in France in 2012 (KR027892 and KR027904). In this study, there was a cluster (n=54) consisted of the HEV-4 sequences being separated in the phylogenetic tree, which could not be classified into an existing subtype, or having inconsistent phylogenetic findings using different methods or reference strains (Supplementary Table 2).

In China's mainland, the most common subtype was 4a (33.7%), whereas it was 4c (32.1%) in Japan and 4b (55.4%) in Hong Kong. Subtype 4b (77.5%) was predominant in Europe, particularly in France (Table 1). Globally, the prevalent subtypes have changed remarkably over the last 18 years. In total, subtypes 4a, 4b, 4d and 4h were most commonly isolated. In 2005-2010, subtype 4a represented 36.9% of isolated HEV-4 sequences; however, it decreased to 9.1% in 2012-2017. Concurrently, the

proportion of subtype 4h sequences increased from 10.1% to 26.3%, and the proportion of subtype 4d sequences increased from 16.6% to 33.6%. Subtype 4h was firstly identified in 2000 and then surpassed 4a in the number of isolation in 2002. In 2018, subtype 4a increased dramatically while subtype 4h decreased (Fig 4).

Phylogenetic analysis of HEV-4

A phylogenetic analysis was performed based on the Met domain (279 nt) and RdRp domain (307 nt) in ORF1, and three ORF2 partial fragments (317 nt, 150 nt, and 537 nt). The ORF2 150 nt partial sequences shared highest nucleotide similarities, whereas the two ORF1 partial sequences demonstrated lower similarities within and between host groups (human and swine) and country groups (China's mainland and Japan) (Table 2). Globally, nucleotide similarities increased by year, particularly between swine sequences on the Met 279 nt, RdRp 307 nt, ORF2 537 nt and human sequences on the ORF2 150 nt (Table 3). The similarities between human sequences isolated in Japan were much higher than those in China's mainland (Table 4). Furthermore, HEV-4 strains isolated in same country shared higher similarities, regardless of host (Table 4).

High nucleotide similarities were observed between countries and regions, such as between France and Denmark (sharing the identity of 96.0%-99.9% on the ORF2 150 nt), South Korea and Mongolia (87.6%-96.7% on the ORF2 150 nt), Hong Kong and Cambodia (92.9%-96.3% on the ORF2 537 nt, which was similar to 94.0%-96.7% within Hong Kong). In addition, 4h isolated in rhesus monkey (n=1), tree shrew (n=1), and wild yak (n=1) shared high similarities based on the two ORF1 regions (96.1%-99.0% on the Met 279 nt, 96.1%-99.0% on the RdRp 307 nt). Similarly, they shared the highest nucleotide similarities (98.8%-99.5%) on the ORF2 317 nt (rhesus monkey, n=11; tree shrew, n=1; wild yak, n=1), whereas 4h sequences in cow (n=52) and goat (n=23) shared high similarities (98.2%-99.0%) on the ORF2 317 nt.

In addition, nucleotide similarities were compared within subtype 4a, 4b, 4d, 4h, and 4i. High similarities were identified between the sequences: 4a, in human and swine on the ORF2 150 nt (93.9%-96.3%); 4b, in Viet Nam and Hong Kong on the Met 279 nt (91.8%-95.0%), in France and Denmark on the ORF2 317 nt (95.5%-97.8%); 4d, in cow and sheep on the ORF2 317 nt (95.6%-

97.8%); 4h, in cow and goat on all the five partial fragments (≥95.9%); 4i, in Japan and Malaysia on the three ORF2 regions (96.1%-98.1% on the ORF2 317 nt, 95.6%-98.5% on the ORF2 150 nt, and 95.3%-97.1% on the ORF2 537 nt).

Discussion

HEV-4 is mainly isolated in Asia and Europe (Dalton & Izopet, 2018b; M. S. Khuroo, Khuroo, & Khuroo, 2016). In this study, we found that the majority of HEV-4 sequences were isolated in China's mainland, Hong Kong, Japan, and France. The geographical distribution was characterized by central distribution in China and France, and circular distribution in the neighboring countries and regions. High nucleotide similarities were observed between the HEV-4 sequences in France and Denmark, Hong Kong and Cambodia, South Korea and Mongolia, suggesting potential cross-border transmission in neighboring areas. Previous studies have provided possible evidence of cross-border transmission, such as transmission of HEV-4 from Southeastern Asia to Europe by international travelers (Fogeda, Avellón, Cilla, & Echevarría, 2009a). In addition, importation pig blood products from China to Europe as feed growth promoters may have contributed to the sporadic human cases of HEV-4 infection in Europe; however, it remains lack of direct evidence (Dalton & Izopet, 2018a). On the other hand, predominant subtypes of HEV-4 varied by area, such as subtype 4a in China's mainland, 4b in Hong Kong and Europe, and 4c in Japan, suggesting that HEV-4 remains endemic. Nucleotide similarities remained high within same countries and regions, regardless of host, as documented elsewhere (Nishizawa et al., 2003). Moreover, the proportion of isolated HEV-4 subtypes has changed over time. Subtype 4a has decreased, whereas 4h and 4d has increased in 2007-2017. On the one hand, the changing prevalence and alternating predominant HEV-4 subtypes might be partly explained by sampling bias, on the other hand, as hypotheses, it may be attributable to increasing human travel and animal trade regionally and globally (Dalton & Izopet, 2018b; Geng et al., 2010). The global distribution of HEV-4 is rapidly changing and emerging in different countries, suggesting the need for more widespread intervention strategies in the future. An HEV vaccine is available (Wu, Chen, Lin, Hao, & Liang, 2016), although it is currently on the market only in China. However, the uptake rate of the vaccine remained low and little-known in general population or public health practitioners (Ren et al., 2019).

HEV-4 subtypes have been increasing in recent years, which are characterized by temporal, spatial, and host species clustering. Subtype 4a has been isolated annually; however, the years of isolation of some subtypes (4c, 4e, 4f, 4g and 4i) remain limited, which demonstrates a higher disease burden in China. So far, swine remains the principal reservoir of HEV-4. Furthermore, a growing number of mammals and non-mammal animals have been documented to carry HEV-4. In this study, we summarized the range of hosts, including swine, wild boar, deer, cattle, goat, sheep, donkey, tiger, leopard, bear, monkey, tree shrew, wild yak, birds, and shellfish. We identified more hosts compared to a previous review in 2016 (Y. Nan & Zhang, 2016). More important, HEV-4 infection in birds (Balayart et al., 1983) may suggest an avian-to-mammalian transmission and that in shellfish (Mohammed Sultan Khuroo, 1980) may suggest a cohabitation in an aquatic environment. It has raised a new public health concern of multiple cross-species transmission, though there was possibility that these hosts were merely opportunistic infection with no proliferation of HEV. Subtypes 4b, 4d, and 4h, which have a variety of animal hosts, were the predominant subtypes in some Asian and European countries and regions; particularly, subtype 4h had increased in 2012-2017 by almost 3 times, suggesting possibly growing cross-species transmission. In addition, subtype 4h sequences showed high nucleotide similarities in wild mammals and domestic animals, respectively, whereas lower similarities between wild mammals and domestic animals, implying continual spread within similar animal hosts. However, it remained unclear if there is animal-to-animal transmission and animal-to-human transmission (D. B. Smith & P. Simmonds, 2018), which warrants further study.

Our study has limitations. Firstly, our study was conducted based on the sequences available in the GenBank, which may represent a sample of HEV-4 strains circulating across the world. Sequencing is associated with multiple factors, such as awareness, research priority, and economic development in every single country and region. In low- and middle-income countries and regions, intention to detect and sequence HEV may be limited, compared to other infectious diseases that may be more important, resulting in sampling and sequencing bias. However, the GenBank is the principal platform for microbial sequences, in which HEV-4 full-length and partial sequences available may be appropriate for understanding the real epidemic scenario of HEV-4, as we could not obtain a more representative sample. Moreover, this study provides a preliminary and quantitative evidence of changing trends in HEV-4 isolation worldwide. Second, that the phylogenetic analysis was performed based on the five

partial fragments within HEV ORF1 and ORF2. Due to all the fragments not parallel available for the same HEV-4 strains and their inconsistent number of isolations by year/country/host, we calculated the nucleotide similarities based on a pooled database. Our analysis with a global perspective may bridge the gap in the absence of each fragment. Third, we identified an uncertain cluster within HEV-4. Some of the sequences were similar to that of subtype 4; however, the sequences could not be currently classified into existing HEV-4 subtypes. It may suggest a new subtype, which should be confirmed with further evidence. At the same time, it showed a limitation in the identification of HEV-4 subtypes based on a short partial fragment (62.9% of the sequences have less than 300 nt in length), as suggested by inconsistent findings in the phylogenetic analysis. Thus, further amplification of longer partial fragment is warranted to determine their subtypes.

In conclusion, China and Japan remain endemic for HEV-4. The number of strains isolated in France may indicate endemicity and highlights the spread of HEV-4 in Europe. HEV-4 subtypes are characterized by temporal, spatial, and host species clustering. Increases in affected areas and animal hosts imply consistent cross-border and cross-species transmission.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 81973101). Dr. Wagner received funding through a Fulbright Scholarship.

References

Asma, J., Stephanie, H. B., Eric, M., Camelia, M., & Anne-Marie, R.-A. (2013). Genotype 4 Hepatitis E Virus in France: An Autochthonous Infection With a More Severe Presentation. *Clinical Infectious Diseases An Official Publication of the Infectious Diseases Society of America, 57*(4), 122-126. doi:10.1093/cid/cit291

Balayart, M. S., Andjaparidze, A. G., Savinskaya, S. S., Ketiladze, E. S., Braginsky, D.

M., Savinov, A. P., & Poleschuk, V. F. (1983). Evidence for a Virus in Non-A, Non-B Hepatitis Transmitted via the Fecal-Oral Route. *Intervirology, 20*(1), 23-31. doi:10.1159/000149370

Colson, P., Romanet, P., Moal, V., Borentain, P., Purgus, R., Benezech, A., . . . Gérolami, R. (2012). Autochthonous Infections with Hepatitis E Virus Genotype 4, France. *Emerging infectious diseases, 18*(8), 1361-1364. doi:10.3201/eid1808.111827

Dalton, H. R., & Izopet, J. (2018a). Transmission and Epidemiology of Hepatitis E Virus Genotype 3 and 4 Infections. *Cold Spring Harb Perspect Med, 8*(11). doi:10.1101/cshperspect.a032144

Dalton, H. R., & Izopet, J. (2018b). Transmission and Epidemiology of Hepatitis E Virus Genotype 3 and 4 Infections. *Cold Spring Harbor Perspectives in Medicine, 8*(11), 20. doi:10.1101/cshperspect.a032144

EASL Clinical Practice Guidelines on hepatitis E virus infection. (2018). *J Hepatol, 68*(6), 1256-1271. doi:10.1016/j.jhep.2018.03.005

Fogeda, M., Avellón, A., Cilla, C. G., & Echevarría, J. M. (2009a). Imported and autochthonous hepatitis E virus strains in Spain. *J Med Virol, 81*(10), 1743-1749. doi:10.1002/jmv.21564

Fogeda, M., Avellón, A., Cilla, C. G., & Echevarría, J. M. (2009b). Imported and autochthonous hepatitis E virus strains in Spain. *Journal of Medical Virology, 81*(10), 1743 - 1749. doi:10.1002/jmv.21564

Garbuglia, A. R., Scognamiglio, P., Petrosillo, N., Mastroianni, C. M., Sordillo, P., Gentile, D., . . . Capobianchi, M. R. (2013). Hepatitis E Virus Genotype 4 Outbreak, Italy, 2011. *Emerging infectious diseases, 19*(1), 110-114. doi:10.3201/eid1901.120983

Geng, Y., Wang, C., Zhao, C., Yu, X., Harrison, T. J., Tian, K., & Wang, Y. (2010). Serological prevalence of hepatitis E virus in domestic animals and diversity of genotype 4 hepatitis E virus in China. *Vector borne and zoonotic diseases, 10*(8), 765-770. doi:10.1089/vbz.2009.0168

Graff, J., Torian, U., Nguyen, H., & Emerson, S. U. (2006). A Bicistronic Subgenomic mRNA Encodes both the ORF2 and ORF3 Proteins of Hepatitis E Virus. *Journal of Virology,*

80(12), 5919-5926. doi:10.1128/JVI.00046-06

Hakze-van der Honing, R. W., van Coillie, E., Antonis, A. F., & van der Poel, W. H. (2011). First isolation of hepatitis E virus genotype 4 in Europe through swine surveillance in the Netherlands and Belgium. *PLoS One, 6*(8), e22673. doi:10.1371/journal.pone.0022673

Khuroo, M. S. (1980). Study of an epidemic of non-A, non-B hepatitis.Possibility of Another Human Hepatitis Virus Distinct From Post-Transfusion non-A, non-B Type. *American Journal of Medicine, 68*(6), 818-824. doi:10.1016/0002-9343(80)90200-4

Khuroo, M. S., Khuroo, M. S., & Khuroo, N. S. (2016). Hepatitis E: Discovery, global impact, control and cure. *World J Gastroenterol, 22*(31), 7030-7045. doi:10.3748/wjg.v22.i31.7030

Kmush, B., Wierzba, T., Krain, L., Nelson, K., & Labrique, A. B. (2013). Epidemiology of hepatitis E in low- and middle-income countries of Asia and Africa. *Semin Liver Dis, 33*(1), 15-29. doi:10.1055/s-0033-1338111

Lu, L., Li, C., & Hagedorn, C. H. (2006). Phylogenetic analysis of global hepatitis E virus sequences: genetic diversity, subtypes and zoonosis. *Rev Med Virol, 16*(1), 5-36. doi:10.1002/rmv.482

Midgley, S., Vestergaard, H. T., Dalgaard, C., Enggaard, L., & Fischer, T. K. (2014). Hepatitis E virus genotype 4, Denmark, 2012. *Emerg Infect Dis, 20*(1), 156-157. doi:10.3201/eid2001.130600

Nan, Y., Wu, C., Qin, Z., & Zhou, E. M. (2017). Zoonotic Hepatitis E Virus: An Ignored Risk for Public Health. *Frontiers in Microbiology, 8*, 2396. doi:10.3389/fmicb.2017.02396

Nan, Y., & Zhang, Y. J. (2016). Molecular Biology and Infection of Hepatitis E Virus. *Frontiers in Microbiology, 7*, 1419. doi:10.3389/fmicb.2016.01419

Nishizawa, T., Takahashi, M., Mizuo, H., Miyajima, H., Gotanda, Y., & Okamoto, H. (2003). Characterization of Japanese swine and human hepatitis E virus isolates of genotype IV with 99 % identity over the entire genome. *The Journal of general virology, 84*(Pt 5), 1245-1251. doi:10.1099/vir.0.19052-0

Ole, W., Sven, S., Judith, K., Martin, K., Camilla, R., Annelie, P., . . . Klaus, S.

(2008). Phylogenetic and Case-Control Study on Hepatitis E Virus Infection in Germany. *Journal of Infectious Diseases, 198*, 1732 - 1741. doi:10.1086/593211

Panda, S. K., & Varma, S. P. K. (2013). Hepatitis E: Molecular Virology and Pathogenesis. *Journal of Clinical & Experimental Hepatology, 3*(2), 114-124. doi:10.1016/j.jceh.2013.05.001

Primadharsini, P. P., Nagashima, S., & Okamoto, H. (2019). Genetic Variability and Evolution of Hepatitis E Virus. *Viruses, 11*(5). doi:10.3390/v11050456

Purdy, M. A., Harrison, T. J., Jameel, S., Meng, X. J., Okamoto, H., Van der Poel, W. H. M., . . . Ictv Report, C. (2017). ICTV Virus Taxonomy Profile: Hepeviridae. *J Gen Virol, 98*(11), 2645-2646. doi:10.1099/jgv.0.000940

Rein, D. B., Stevens, G. A., Theaker, J., Wittenborn, J. S., & Wiersma, S. T. (2012). The global burden of hepatitis E virus genotypes 1 and 2 in 2005. *Hepatology, 55*(4), 988-997. doi:10.1002/hep.25505

Ren, H., Wagner, A. L., Xie, J. Y., Chen, K. Y., Lu, Y. H., Zheng, X. B., . . . Chen, X. X. (2019). How Do Experts and Nonexperts Want to Promote Vaccines? Hepatitis E Vaccine as Example. *Health Serv Insights, 12*, 1178632919897276. doi:10.1177/1178632919897276

Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *J Virol, 84*(19), 9733-9748. doi:10.1128/jvi.00694-10

Smith, D. B., Izopet, J., Nicot, F., Simmonds, P., Jameel, S., Meng, X. J., . . . Purdy, M. A. (2020). Update: proposed reference sequences for subtypes of hepatitis E virus (species Orthohepevirus A). *J Gen Virol, 101*(7), 692-698. doi:10.1099/jgv.0.001435

Smith, D. B., & Simmonds, P. (2018). Classification and Genomic Diversity of Enterically Transmitted Hepatitis Viruses. *Cold Spring Harb Perspect Med, 8*, a031880. doi:10.1101/cshperspect.a031880

Smith, D. B., & Simmonds, P. (2018). Classification and Genomic Diversity of Enterically Transmitted Hepatitis Viruses. *Cold Spring Harb Perspect Med, 8*(9). doi:10.1101/cshperspect.a031880

Smith, D. B., Simmonds, P., Izopet, J., Oliveira-Filho, E. F., & Purdy, M. A. (2016).

Proposed reference sequences for Hepatitis E virus subtypes. *Journal of General Virology, 97*(3), 537. doi:10.1099/jgv.0.000393

Smith, D. B., Simmonds, P., Jameel, S., Emerson, S. U., Harrison, T. J., Meng, X. J., . . . Purdy, M. A. (2015). Consensus proposals for classification of the family Hepeviridae. *Journal of General Virology, 96*(Pt_10), 2223-2232. doi:10.1099/vir.0.068429-0

Smith, D. B., Simmonds, P., Members Of The International Committee On The Taxonomy Of Viruses Hepeviridae Study, G., Jameel, S., Emerson, S. U., Harrison, T. J., . . . Purdy, M. A. (2015). Consensus proposals for classification of the family Hepeviridae. *J Gen Virol, 96*(Pt 5), 1191-1192. doi:10.1099/vir.0.000115

Sridhar, S., Yip, C., Wu, S., Cai, J., Zhang, J., Leung, K., . . . Teng, L. (2018). Rat Hepatitis E Virus as Cause of Persistent Hepatitis after Liver Transplant. *Emerging infectious diseases, 24*(12), 2241-2250. doi:10.3201/eid2412.180937

Tam, A. W., Smith, M. M., Guerra, M. E., Huang, C.-C., Bradley, D. W., Fry, K. E., & Reyes, G. R. (1991). Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. *Transboundary and emerging diseases, 185*(1), 120-131. doi:10.1016/0042-6822(91)90760-9

Tessé, S., Lioure, B., Fornecker, L., Wendling, M.-J., Stoll-Keller, F. o., Bigaillon, C., & Nicand, E. (2012). Circulation of genotype 4 hepatitis E virus in Europe: First autochthonous hepatitis E infection in France. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology, 54*(2), 197-200. doi:10.1016/j.jcv.2012.02.007

Tsarev, S. A., Emerson, S. U., Reyes, G. R., Tsareva, T. S., Letgers, L. J., Malik, I. A., . . . Purcell, R. H. (1992). Characterization of a Prototype Strain of Hepatitis E Virus. *Proceedings of the National Academy of Sciences of the United States of America, 89*(2), 559-563. doi:10.1073/pnas.89.2.559

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. doi:10.1016/j.ebiom.2016.07.039

Wang, Y., Ling, R., Erker, J. C., Zhang, H., Li, H., Desai, S., . . . Harrison, T. J.

(1999). A divergent genotype of hepatitis E virus in Chinese patients with acute hepatitis. *The Journal of general virology, 80*, 169-177. doi:10.1099/0022-1317-80-1-169

WHO. (2010). Global prevalence of hepatitis E: a systematic review. Available at: http://whqlibdoc.who.int/hq/2010/WHO_IVB_10.14_eng.pdf.

WHO. (2019). Hepatitis E Available at: https://www.who.int/news-room/fact-sheets/detail/hepatitis-e.

Wu, X., Chen, P., Lin, H., Hao, X., & Liang, Z. (2016). Hepatitis E virus: Current epidemiology and vaccine. *Hum Vaccin Immunother, 12*(10), 2603-2610. doi:10.1080/21645515.2016.1184806

Zhang, S., Wang, J., Yuan, Q., Ge, S., Zhang, J., Xia, N., & Tian, D. (2011). Clinical characteristics and risk factors of sporadic Hepatitis E in central China. *Virology Journal, 8*(1), 152. doi:10.1186/1743-422X-8-152

Table 1 Geographical distribution of HEV-4 strains (n=2295)

Countries and regions	Number of sequences	Predominant subtype/cluster	Other existing subtypes/clusters
Asia			
China's mainland	1745	4a (33.7%)	4, 4b, 4c, 4d, 4f, 4g, 4h, 4i, uncertain
Japan	243	4c (32.1%)	4, 4a, 4b, 4d, 4e, 4f, 4g, 4h, 4i, uncertain
Hong Kong, China	92	4b (55.4%)	4, 4a, 4d, 4f, 4h, 4i, uncertain
South Korea	37	4a (70.3%)	4c,4d, uncertain
India	27	4e (88.9%)	4b
Taiwan, China	21	4c (57.1%)	4, 4a, 4b, uncertain
Laos	13	4g (38.5%)	4, 4a, 4f, uncertain
Viet Nam	15	4b (73.3%)	4a, 4c, 4h
Mongolia	12	4a (100%)	-

Indonesia	9	4b (77.8%)	4i, uncertain
Malaysia	8	4i (75.0%)	4a, 4g
Cambodia	1	4b	-
Subtotal	2223		
Europe			
France	48	4b (91.7%)	4, 4d, 4h
Russia	9	4b (88.9%)	4
Denmark	3	4b (66.7%)	4d
Italy	3	4d (100%)	-
Belgium	3	uncertain (100%)	-
Switzerland	2	4g	-
Spain	1	4b	-
Germany	1	4	-
United Kingdom	1	4e	-
Subtotal	71		
America			
USA	1	4a	-

* Cluster uncertain consisted of the HEV-4 sequences being separated in the phylogenetic tree, which could not be classified into an existing subtype, or was composed of the HEV-4 sequences that had inconsistent phylogenetic findings using neighbor-joining and maximum methods or 48 full-length HEV reference strains and 10 HEV-4 reference

Table 2 Nucleotide identities on the five partial fragments stratified by principal hosts and countries (SE)

	Met 279 nt	RdRp 307 nt	ORF2 317 nt	ORF2 150 nt	ORF2 537 nt
Hosts					
within human	85.9 (1.6)	83.7 (1.6)	86.4 (1.3)	90.5 (1.5)	86.3 (1.1)

	n=177	n=289	n=304	n=208	n=276
within swine	86.5 (1.6)	85.3 (1.5)	86.2 (1.4)	88.7 (1.8)	85.6 (1.1)
	n=32	n=141	n=243	n=126	n=33
between human & swine	84.6 (1.7)	84.0 (1.6)	85.0 (1.5)	89.0 (1.8)	85.8 (1.1)
Countries					
within China's mainland	86.5 (1.5)	84.7 (1.5)	86.1 (1.4)	89.5 (1.7)	86.7 (1.1)
	n=55	n=375	n=482	n=311	n=266
within Japan	88.9 (1.3)	87.8 (1.3)	89.9 (1.1)	92.3 (1.5)	90.0 (0.9)
	n=111	n=56	n=181	n=393	n=38
between China's mainland & Japan	83.6 (1.8)	81.5 (1.9)	84.95 (1.5)	88.6 (1.9)	84.4 (1.2)

Table 3 Nucleotide identities on the five partial fragments stratified by principal hosts and years (SE)

	Met 279 nt	RdRp 307 nt	ORF2 317 nt	ORF2 150 nt	ORF2 537 nt
Before 2000,	88.1 (1.5)	87.8 (1.4)	87.8 (1.3)	92.9 (1.4)	90.7 (0.8)
Human	n=20	n=19	n=29	n=14	n=14
2001-2005,	84.8 (2.6)	NA	87.2 (1.5)	92.1 (1.5)	NA
Swine	n=2		n=35	n=17	
2001-2005,	87.3 (1.5)	84.8 (1.6)	89.0 (1.2)	91.8 (1.4)	86.7 (1.1)
Human	n=68	n=46	n=75	n=63	n=44
2006-2010,	85.7 (1.7)	84.5 (1.6)	87.9 (1.3)	89.6 (1.6)	85.1 (1.2)
Swine	n=14	n=87	n=79	n=67	n=14
2006-2010,	87.3 (1.5)	84.5 (1.6)	90.2 (0.9)	91.0 (1.4)	86.9 (1.0)
Human	n=55	n=185	n=56	n=95	n=174
2011-2015,	88.0 (1.6)	87.3 (1.3)	89.3 (1.1)	93.2 (1.3)	86.4 (1.1)
Swine	n=7	n=44	n=81	n=20	n=9
2011-2015,	86.8 (1.6)	83.5 (1.7)	85.7 (1.4)	89.6 (1.8)	85.5 (1.1)

Human	n=31	n=26	n=90	n=21	n=41
2016-2019,	93.7 (0.9)	91.5 (1.1)	87.1 (1.4)	91.6 (1.7)	92.2 (0.7)
Swine	n=9	n=9	n=48	n=22	n=9
2016-2019,	88.2 (2.0)	86.2 (1.6)	89.9 (1.1)	94.1 (1.3)	89.3 (0.7)
Human	n=3	n=13	n=54	n=15	n=3

* NA: there was a single sequence.

Table 4 Nucleotide identity on the five partial fragments stratified by combined principal hosts and countries (SE)

	Met 279 nt	RdRp 307 nt	ORF2 317 nt	ORF2 150 nt	ORF2 537 nt
Within the group					
Japan, human	89.7 (1.2)	87.95 (1.4)	91.3 (1.0)	92.4 (1.5)	90.1 (0.9)
	n=100	n=54	n=141	n=37	n=36
Japan, swine	84.8 (2.5)	84.1 (2.5)	92.6 (1.1)	87.2 (3.3)	87.8 (1.6)
	n=2	n=2	n=16	n=2	n=2
China's mainland,	86.1 (1.6)	84.7 (1.6)	86.4 (1.4)	91.2 (1.5)	87.1 (1.0)
human	n=14	n=223	n=106	n=156	n=221
China's mainland,	86.7 (1.6)	85.5 (1.5)	86.5 (1.4)	89.0 (1.8)	86.0 (1.1)
swine	n=27	n=137	n=218	n=122	n=30
Between the groups					
Japan, human & swine	89.0 (1.4)	85.8 (1.7)	89.4 (1.4)	91.1 (1.9)	89.3 (1.0)
China's mainland, human & swine	85.3 (1.6)	84.8 (1.5)	84.9 (1.5)	89.3 (1.7)	86.4 (1.1)
Human, Japan & China's mainland	83.6 (1.8)	81.4 (1.9)	84.9 (1.6)	89.2 (1.9)	84.4 (1.3)
Swine, Japan & China's mainland	84.1 (1.9)	82.3 (1.9)	84.5 (1.7)	88.3 (2.0)	83.4 (1.4)

Japan, swine & China's mainland, human	84.6 (1.8)	83.3 (1.8)	85.0 (1.6)	89.6 (1.9)	83.6 (1.4)
Japan, human & China's mainland, swine	83.3 (1.9)	81.1 (2.0)	84.7 (1.6)	88.3 (1.9)	84.3 (1.3)

* NA: there was a single sequence.

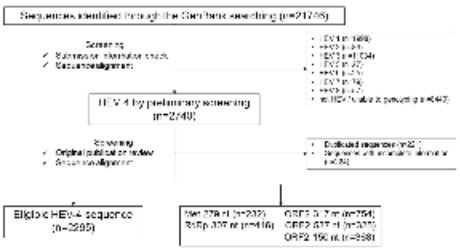
Figure legends

Fig 1 The roadmap of the inclusion of hepatitis E virus genotype 4 (HEV-4) sequences and phylogenetic analysis in the study.

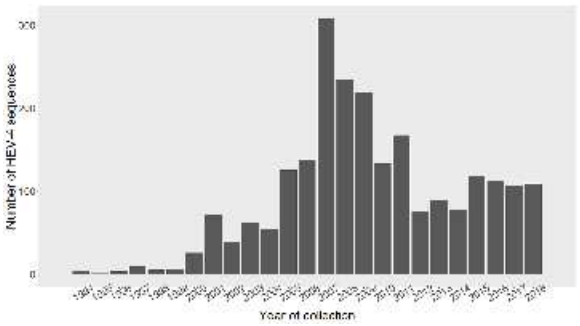
Fig 2 Temporal distribution of isolated hepatitis E virus genotype 4 (HEV-4) strains in 1992-2018.

Fig 3 Geographical distribution of isolated hepatitis E virus genotype 4 (HEV-4) strains.

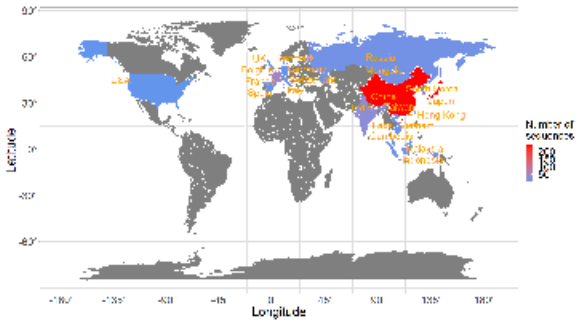
Fig 4 Isolation of hepatitis E virus genotype 4 (HEV-4) subtypes in 1992-2018.



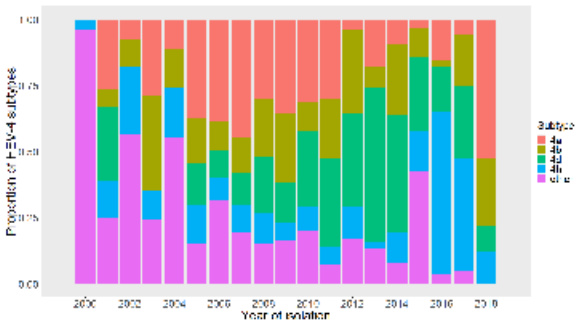
zph_12934_f1.tif



zph_12934_f2.tif



zph_12934_f3.tif



zph_12934_f4.tif