Infection, Genetics and Evolution 43 (2016) 267–273

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

New insights into the hepatitis E virus genotype 3 phylodynamics and evolutionary history

Santiago Mirazo [a,1], Daiana Mir [b,1], Gonzalo Bello [b], Natalia Ramos [a], Héctor Musto [c], Juan Arbiza [a,*]

[a] Sección Virología, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay
[b] Laboratorio de AIDS & Imunologia Molecular, Instituto Oswaldo Cruz. Av. Brasil 4365, 21045900 Rio de Janeiro, Brazil
[c] Laboratorio de Organización y Evolución del Genoma, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400. Montevideo, Uruguay

ARTICLE INFO

Article history:
Received 13 February 2016
Received in revised form 19 May 2016
Accepted 1 June 2016
Available online 3 June 2016

Keywords:
Hepatitis E virus genotype 3 phylodynamics
South America
Evolutionary history

ABSTRACT

Hepatitis E virus (HEV) is an emergent hepatotropic virus endemic mainly in Asia and other developing areas. However, in the last decade it has been increasingly reported in high-income countries. Human infecting HEV strains are currently classified into four genotypes (1–4). Genotype 3 (HEV-3) is the prevalent virus genotype and the mostly associated with autochthonous and sporadic cases of HEV in developed areas. The evolutionary history of HEV worldwide remains largely unknown. In this study we reconstructed the spatiotemporal and population dynamics of HEV-3 at global scale, but with particular emphasis in South America, where case reports have increased dramatically in the last years. To achieve this, we applied a Bayesian coalescent-based approach to a comprehensive data set comprising 97 GenBank HEV-3 sequences for which the location and sampling date was documented. Our phylogenetic analyses suggest that the worldwide genetic diversity of HEV-3 can be grouped into two main Clades (I and II) with a T_{mrca} dated in approximately 320 years ago (95% HPD: 420–236 years) and that a unique independent introduction of HEV-3 seems to have occurred in Uruguay, where most of the human HEV cases in South America have been described. The phylodynamic inference indicates that the population size of this virus suffered substantial temporal variations after the second half of the 20th century. In this sense and conversely to what is postulated to date, we suggest that the worldwide effective population size of HEV-3 is not decreasing and that frequently sources of error in its estimates stem from assumptions that the analyzed sequences are derived from a single panmictic population.
Novel insights on the global population dynamics of HEV are given. Additionally, this work constitutes an attempt to further describe in a Bayesian coalescent framework, the phylodynamics and evolutionary history of HEV-3 in the South American region.

1. Introduction

Hepatitis E virus (HEV) is the aetiological agent of acute hepatitis E; an infection considered endemic in many countries from Africa and Asia (Purcell and Emerson, 2001). HEV is transmitted primarily by fecal–oral routes, and has been reported to occur as large waterborne epidemics and small outbreaks in developing areas (Dawson et al., 1992; Donati et al., 1997). It has been estimated that two billion people, representing one third of the world population live in endemic areas for HEV and, therefore, are at risk of infection (Holla et al., 2013). However, over the last decade, it has occurred in several high income countries an increasing number of sporadic locally acquired cases in which it is often not possible to establish the route of acquisition of infection (Echevarría et al., 2013). Additionally, accumulating lines of evidence indicates that hepatitis E is a zoonotic infection with pigs and wild boars serving as the main reservoir for human infections (Meng, 2011; Kumar et al., 2013). The global burden of HEV infection is thought to be due to sporadically transmitted cases rather than hepatitis E outbreaks (Kumar et al., 2013).

Human infecting HEV sequences have been classified into four major recognized genotypes (1–4), according to analyses of the complete genome sequences and/or variable partial HEV genomic regions within the ORF1 and ORF2 (Schlauder and Mushahwar, 2001; Lu et al., 2006). Genotypes 1 and 2 (HEV-1 and 2) were detected in Asia, Africa and Latin America and are mostly associated with large epidemics and outbreaks (Purcell and Emerson, 2001; Okamoto, 2007; Teo, 2010). HEV-3 and 4, by contrast, comprise zoonotic strains (Okamoto et al., 2001; Tei et al., 2003; Sonoda et al., 2004; Sato et al., 2011) and give rise to mainly autochthonous HEV infections in non-endemic high-income countries (Meng, 2011). HEV-3 is worldwide distributed and has been proposed as the most prevalent HEV clade circulating in Latin America (Mirazo et al., 2013), whereas HEV-4 has been reported in countries from Asia

* Corresponding author.
E-mail address: jarbiza@fcien.edu.uy (J. Arbiza).
[1] These authors contribute equally to this work.

http://dx.doi.org/10.1016/j.meegid.2016.06.003
1567-1348/

and Central Europe (Nishizawa et al., 2003; Wichmanm et al., 2008; Hakze-van der Honing et al., 2011; Tessé et al., 2012; Pischke et al., 2014).

There is little information about the evolutionary and demographic history of HEV-3 in human and swine populations. Few attempts to shed light on this issue have been carried out at both regional and global levels (Purdy and Khudyakov, 2010; Nakano et al., 2012a; Purdy et al., 2012; Zehender et al., 2014). In a global context, the reported studies agree that HEV-3 infections remained relatively constant until the turn of the 20th century, and then experienced a dramatic increase between the 1940 and 1965, reaching then a plateau, followed by a rapid decline in the recent years (Purdy and Khudyakov, 2010; Nakano et al., 2012a; Zehender et al., 2014). Increased consumption of pork meat and pork farming activities has been consistently pointed as the most probable causative factor of HEV-3 population growth during the last century; but the possible causes of HEV-3 population decline towards the present remain unclear. One study suggests that the decrease in the effective population size of HEV-3 in Japan could be due either to an increase health surveillance to prevent swine-specific diseases or to an artifact related to non-random sampling of analyzed sequences (Nakano et al., 2012a, 2012b).

However, at regional scale, so far, there is no information concerning the origin and genetic diversity of HEV-3 in South America, as the solely study performed up to date was limited to a local community from Southern Bolivia (Purdy et al., 2012).

We have recently reported a complete phylogenetic analysis of a set of Uruguayan strains, in which all samples seemed to be very closely related to European HEV strains, and were quite dissimilar to South American isolates (Mirazo et al., 2013). Here, we move forward, in an attempt to better understand the complex molecular epidemiology and phylodynamics of HEV in this developing region.

This work aims two goals. The first one consists in the characterization of the major global HEV-3 lineages in order to obtain evidences regarding the origin of HEV-3 strains circulating worldwide, and particularly in South America, by analysing all available sequences that matched the methodological criteria. The second one attempts to infer the globally HEV-3 demographic history to test whether the signal of global population decline previously described was or was not associated to the sampling of epidemiologically linked sequences from a non-panmictic population. To reach these goals we applied Maximum Likelihood and Bayesian coalescent-based frameworks to analyse 97 HEV-3 ORF2 sequences isolated worldwide over a period of 22 years (1993 to 2015).

2. Materials and methods

2.1. Sequence dataset

All HEV-3 ORF2 sequences with known sampling date and location available at the GenBank by April 2016 were included. This resulted in a final dataset of 97 HEV ORF2 sequences from 15 different countries, with sampling dates ranging from 1993 to 2015. GenBank accession number, location and year of isolation of sequences included are shown in Table S1. Nucleotide sequences were aligned using MAFFT v6.902b program (Katoh et al., 2002), resulting in an alignment of 756 nucleotides length (nt 5294 to 6050, in the US1 reference strain). Alignment is available from the authors upon request.

2.2. Subsampling of epidemiologically linked sequences

To take into account the potential relationships between viral population subdivision (given by the use of closely related sequences belonging to the same HEV-3 outbreak) and the decline in viral population size (Heller et al., 2013), we generated several subsets by removing closely related sequences within single outbreaks. To do this, sequences were grouped by similarity with the CD-HIT program (Li and Godzik, 2006) and only one sequence per cluster was selected. Clustering by CD-HIT was made at 100%, 0.99% and 0.98% identity levels.

2.3. Phylogenetic tree reconstruction

To investigate the phylogenetic relationships, a maximum likelihood tree was constructed with the PhyML 3.0 program under the GTR + I + Γ4 model of nucleotide substitution as determined by ModelGenerator v0.851. The heuristic tree search was performed using the SPR branch-swapping algorithm and branch support was calculated with the approximate likelihood-ratio (aLRT) SH-like test (Anisimova and Gascuel, 2006).

2.4. Coalescent analyses

The evolutionary rate (μ, nucleotide substitutions per site per year, subs./site/year), the age of the most recent common ancestor (T_{mrca}, years), the ancestral geographic movements, and the demographic pattern were jointly estimated, using the Bayesian Markov Chain Monte Carlo (MCMC) statistical framework implemented in BEAST v1.8.0 package (Drummond and Rambaut, 2007; Drummond et al., 2012). Coalescent analyses were conducted on the complete dataset and all subsets that resulted from the subsampling at different levels of identity. Analyses were carried out with a piece-wise constant-multiple change process called Bayesian Skyline Plot (BSP) as the coalescent tree prior, under the GTR + I + Γ4 model of nucleotide substitution (Drummond et al., 2005) and a relaxed uncorrelated lognormal molecular clock model (Drummond et al., 2006). A uniform prior was applied on the clock rate ($1.0–3.6 \times 10^{-3}$, initial value: 1.5×10^{-3}, subs./site/year) on the basis of estimations reported from previous studies (Purdy and Khudyakov, 2010; Nakano et al., 2012a; Zehender et al., 2014). Migration events throughout the phylogenetic history were inferred using both a reversible and non-reversible substitution model for discrete traits coupled with a Bayesian stochastic search variable selection (BSSVS) procedure. Model fit was evaluated using (log) marginal likelihood estimates obtained through path sampling (Lartillot and Philippe, 2006) and stepping-stone sampling (Xie et al., 2011). Migratory events and significant nonzero rates obtained by the BSSVS approach were summarized using the cross-platform SPREAD application. MCMC sampling was performed for 200 million generations and convergence of parameters was assessed by calculating the Effective Sample Size (ESS) using TRACER v1.5 program (Drummond and Rambaut, 2007) after excluding the initial 10% of each run. Uncertainty in parameter estimates was reflected in the 95% Highest Probability Density (HPD) values. The maximum clade credibility (MCC) tree was inferred from the posterior distribution of trees using TreeAnnotator v1.8 with a burn-in rate of 10% and visualized with FigTree v1.4. A graphical representation of the effective number of infections through time and the number of lineages-through-time (LTT), were generated with R-statistical software package.

3. Results

3.1. Phylogenetic analysis

Examination of the phylogenetic reconstruction of the 97 HEV-3 ORF2 partial sequences suggests that the global diversity of this genotype could be grouped into two highly supported (aLRT = 1) main clades (I and II) that contained different ratios of Asian, European, African and South/North American strains (Fig. 1). Clade I enclosed 82% of the sequences included in our study and comprehended three highly supported (aLRT ≥ 0.85) sub Clades (IA–IC). The sub Clade IA was almost exclusively made up of Japanese isolates (95%), sampled between 1998 and 2006. This sub clade also included one Chinese and one Canadian sequence isolated in 2008 and 2003. The sub Clade IB comprised a set of sequences from Asia, Europe and North America, sampled between