Virus Research 233 (2017) 60–69

Contents lists available at ScienceDirect

Virus Research

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

Estimation of main diversification time-points of hantaviruses using phylogenetic analyses of complete genomes

Guillaume Castel[a,*], Noël Tordo[b,c], Alexander Plyusnin[d]

[a] INRA-UMR 1062 CBGP, 755 Avenue Campus Agropolis, CS30016, 34988 Montferrier sur Lez, France
[b] Unit Antiviral Strategies, Institut Pasteur, 25 Rue du Dr Roux, 75724 Paris Cedex 15, France
[c] Institut Pasteur de Guinée, Gamal Abdel Nasser University, Conakry, Guinea
[d] Department of Virology, University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 7 November 2016
Received in revised form 8 March 2017
Accepted 9 March 2017
Available online 16 March 2017

Keywords:
Hantavirus
Diversification
Co-speciation
Phylogeny

ABSTRACT

Because of the great variability of their reservoir hosts, hantaviruses are excellent models to evaluate the dynamics of virus-host co-evolution. Intriguing questions remain about the timescale of the diversification events that influenced this evolution. In this paper we attempted to estimate the first ever timing of hantavirus diversification based on thirty five available complete genomes representing five major groups of hantaviruses and the assumption of co-speciation of hantaviruses with their respective mammal hosts. Phylogenetic analyses were used to estimate the main diversification points during hantavirus evolution in mammals while host diversification was mostly estimated from independent calibrators taken from fossil records. Our results support an earlier developed hypothesis of co-speciation of known hantaviruses with their respective mammal hosts and hence a common ancestor for all hantaviruses carried by placental mammals.

1. Introduction

Hantaviruses constitute the genus *Hantavirus* in the family *Bunyaviridae* (Plyusnin et al., 2012). The family includes three other genera of mammalian viruses: *Orthobunyavirus*, *Phlebovirus*, and *Nairovirus*, while the genus *Tospovirus* consists of plants viruses. Hantaviruses are an exception in the family *Bunyaviridae* since they are not transmitted by bites of arthropod vectors, but directly by aerosols from one infected mammal (rodent, bat or insectivore) to another, including human where hantaviruses can provoke Haemorrhagic Fever with Renal Syndrome (HFRS) or Hantavirus Cardio Pulmonary Syndrome (HCPS). However, mosquito-born (MOSBO) hantaviruses have been recently reported (Li et al., 2015), a discovery that might have important consequences and change our view on the hantavirus evolution in general.

The prototype *Hantaan virus* (HTNV), that has given the name to the genus, has been discovered back in the 1980-ies in Korea as a causative agent of a disease previously known as Korean Haemorrhagic Fever, then HFRS [for a review, see (Lee et al., 2014)]. Later, more hantaviruses provoking HFRS in humans have been found in Asia (*Seoul virus*, SEOV (Lee et al., 1982) and others) and Europe (*Puumala virus*, PUUV (Brummer-Korvenkontio et al., 1980), and others). In the Americas hantaviruses were mostly discovered later (LeDuc et al., 1984) and often associated with HCPS (Nichol et al., 1993; Padula et al., 1998). Hantaviruses were thought for several decades to be associated exclusively with rodents (order *Rodentia*) and although the first hantavirus isolated in cell culture, *Thottapalayam virus* (TPMV), originated from a shrew (Song et al., 2007), it was initially believed to represent a spill-over event from a rodent. Consequently, the collective term "ROBO-viruses" (from ROdent-BOrne) has been initially coined. The recent outburst of novel hantavirus genotypes (likely to represent novel species as well) in shrews and moles (order *Eulipotyphla*) and later in bats (order *Chiroptera*) has proven the variety of mammal hosts being much wider than initially anticipated [for a review, see (Yanagihara et al., 2014; Zhang, 2014). These groups of hantaviruses could be called INBO- (from Insectivore-BOrne) and BABO (from Bat-BOrne) hantaviruses, respectively. Most recently, hantavirus partial genome sequences (encoding the L protein) have been recovered from arthropods: Culex and Armigeres mosquitoes (Li et al., 2015). This finding was in line with other publications on arthropod-specific

* Corresponding author.
E-mail addresses: guillaume.castel@supagro.inra.fr (G. Castel), ntordo@pasteur.fr (N. Tordo), Alexander.Plyusnin@Helsinki.Fi (A. Plyusnin).

http://dx.doi.org/10.1016/j.virusres.2017.03.011
0168-1702/

bunya- and togaviruses that are thought to maintain in insects only (reviewed in (Junglen and Drosten, 2013)). These viruses would be logically called "MOSBO-hantaviruses" (from MOSquito-BOrne) (this term is used in the present manuscript awaiting for a possible more appropriate one). In addition to their discovery in new hosts, hantaviruses have continued to expand geographically and plenty of novel genotypes/species were found in Africa, thus enlarging "the hantavirus world" to this continent (for a review, see (Witkowski et al., 2014)).

Although their medical importance was the initial reason to focus on hantaviruses (Jonsson et al., 2010), they rapidly appeared as excellent models to study virus evolution through phylogenetic analyses (Antic et al., 1992; Nichol et al., 1993; Plyusnin et al., 1996). Their genome evolves with relatively low speed and strong stabilizing selection (for a review, see (Holmes and Zhang, 2015; Sironen and Plyusnin, 2011)). The principal mechanism generating genetic diversity is genetic drift, i.e. gradual accumulation of point mutations, mostly neutral or quasi-neutral, and small deletions/insertions in the non-coding regions. Reassortment of genome RNA segments (Bennett et al., 2014) and, to a lesser extent, recombination may have also contributed (Plyusnin et al., 2002).

Although some examples of host switching have been described (Guo et al., 2013; Holmes and Zhang, 2015; Lin et al., 2012; Vapalahti et al., 1996), the main evolutionary forces, natural selection and genetic drift, have been shaping hantavirus diversification in their respective hosts and/or geographic location (Bennett et al., 2014). As hantaviruses have a tri-segmented RNA genome, they also have been subjected to reassortment events. During the two last decades, geographic clustering of genetic variants has been demonstrated, supporting the hypothesis of a close association and hence long-term co-speciation with natural hosts on a timescale of millions of years (MY) (Plyusnin and Morzunov, 2001; Plyusnin and Sironen, 2014; Sironen et al., 2001). This hypothesis has been challenged by several studies evidencing different timescales between viral and host evolution (Ramsden et al., 2009; Souza et al., 2014). However, these studies used short-term rates of evolution not well adapted to estimate very ancient divergence events in phylogenies (Castel et al., 2014; Sharp and Simmonds, 2011). Moreover, the use of common models (Wertheim and Kosakovsky Pond, 2011) with mutation rates calculated on recent viral genes evolving under strong purifying selection like hantavirus (Castel et al., 2014; Hjelle et al., 1995) can lead to severe underestimation of divergences for viral ancestors (Taylor et al., 2014; Wertheim and Kosakovsky Pond, 2011). Today, the gradual accumulation of complete genome sequences in the databases helps to precise the initial estimates based on partial sequences (see (Sironen et al., 2001) for example) and thus to improve our understanding of hantavirus genetics and speed of evolution.

One of the most intriguing question remains the timescale of diversification events in hantavirus evolution (Bennett et al., 2014; Holmes and Zhang, 2015). This point is difficult to address due to the lack of fossils of viruses. To overcome this point, an alternative consists at using fossil records of the host to calibrate the phylogenetic trees (Gustafsson et al., 2010; Ho et al., 2011). Of course, the origin of animals (Metazoa) itself might be viewed as controversial: e.g. the dating based on molecular data suggests around 800 million years, a projection that is older than the fossil evidence of the "Cambrian Explosion" estimated from 550 million years to 100 million years for the most recent estimations (Goswami, 2012). There are other limitations as well. But, those do not seem to make the approach less fruitful.

Based on the assumption that ancient hantaviruses were already present at the points of diversification of major placental clades (Plyusnin and Sironen, 2014) and would have co-evolved with their hosts, we attempt in this paper to estimate the main diversification time-points during hantavirus evolution for the last 100 MY using the accumulated wealth of 35 hantaviruses with completely sequenced genomes. We confront them to the current knowledge of host diversification based on independent calibrators taken from fossil records.

2. Materials and methods

2.1. Sequence datasets

Table 1 displays the list of the selected the 35 complete coding genome sequences of hantaviruses, with the three segments S-M-L that were used either separately or concatenated for phylogenetic studies. They encompass *Murinae*-, *Arvicolinae*-, *Sigmodontinae*-, and *Neotominae*- ROBO- as well as INBO- and BABO- hantaviruses. In addition, the complete L protein-encoding sequence of the recently described MOSBO-hantavirus *Jianxia Mosquito Virus 2* (JMV-2) (Li et al., 2015) was compared to the 35 L protein-encoding sequences mentioned above.

2.2. Phylogenetic analyses

Multiple sequence alignments were prepared with the MUSCLE alignment program (Edgar, 2004) implemented in SEAVIEW v4.5.4 (Gouy et al., 2010). The GBLOCK program (Talavera and Castresana, 2007) was then used to remove gaps and poorly aligned positions to improve the pertinence of alignments for phylogenetic analyses. Phylogenies were inferred using the Maximum Likelihood (ML) approach in PhyML v3.0 (Guindon et al., 2010) (implemented in SEAVIEW v4.5.4.), with a statistical approximate likelihood ratio test (aLRT) of branch support. The Model Test function implemented in MEGA6 (Tamura et al., 2013) allowed to identify the GTR + G + I model (General Time Reversible, with the Gamma-distribution allowing some sites to be evolutionary invariable) as the optimal substitution model. The transition/transversion ratio was fixed to the value 4.0 and nucleotide (nt) frequencies were calculated from the dataset. Rate heterogeneity was applied using the discrete gamma distribution with four rate categories, and the shape parameter alpha was estimated from the dataset.

Phylogenetic trees were compared by calculating the Robinson-Foulds (RF) distance measuring the topological distance between unrooted phylogenetic trees (Robinson and Foulds, 1981) with the CompPhy program (Fiorini et al., 2014) available online on the ATGC bioinformatics platform at ATGC (2017) http://www.atgc-montpellier.fr/compphy/.

2.3. Evolutionary distances

Genetic distances between Hantavirus groups were calculated using functions implemented in the MEGA6 program. Analyses were conducted using the Maximum Composite Likelihood substitution model (for nucleotide—nt sequences) or the Poisson substitution model (for amino acid—aa sequences). The rate variation among sites was modeled with the gamma-distribution (shape parameter alpha = 1). All the other parameters were set to their default values.

2.4. Estimation of the divergence time points

Divergence times for all branching points in the phylogenetic tree were calculated with the RelTime method described by Tamura et al. (2012) and implemented in MEGA6. This method supports multiple user-defined calibration constraints. The Maximum Likelihood method based on the GTR-model was used and relative times were optimized and converted to absolute divergence times