Infection, Genetics and Evolution 20 (2013) 304–311

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

The variability of the large genomic segment of Ťahyňa orthobunyavirus and an all-atom exploration of its anti-viral drug resistance

Patrik Kilian [a,b], James J. Valdes [b], Daniel Lecina-Casas [c], Tomáš Chrudimský [a], Daniel Růžek [a,b,d,*]

[a] Faculty of Science, University of South Bohemia, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
[b] Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
[c] Joint BSC-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, c/Jordi Girona 29, 08034 Barcelona, Spain
[d] Department of Virology, Veterinary Research Institute, Hudcova 70, CZ-62100 Brno, Czech Republic

ARTICLE INFO

Article history:
Received 23 August 2013
Received in revised form 19 September 2013
Accepted 21 September 2013
Available online 30 September 2013

Keywords:
Ťahyňa virus
Orthobunyavirus
California complex
Genetic variability
Large genomic segment

ABSTRACT

Ťahyňa virus (TAHV), a member of the *Bunyaviridae* family (California complex), is an important but neglected human mosquito-borne pathogen. The virus genome is composed of three segments, i.e., small (S), medium (M), and large (L). Previous studies on genetic variability of viruses within the California complex were focused on S and M segments, but the L segment remains relatively unstudied. To assess the genetic variation and the relation to virus phenotype we analyzed the L segment sequences of biologically diverse TAHV strains isolated in the Czech Republic and Slovakia. Phylogenetic analysis covering all available sequences of the L segment of TAHV clearly revealed two distinguished lineages, tentatively named as "European" and "Asian". The L segment strains within the European lineage are highly conserved (identity 99.3%), whilst Asian strains are more genetically diverse (identity 97%). Based on sequence comparison with other bunyaviruses, several non-synonymous nucleotide substitutions unique for TAHV in the L segment were identified. We also identified specific residue substitutions in the endonuclease domain of TAHV compared with the La Crosse virus. Since the endonuclease domain of the La Crosse virus has been resolved, we employed an all energy landscape algorithm to analyze the ligand migration of a viral polymerase inhibitor. This allowed us to demonstrate, at the atomic level, that this viral polymerase inhibitor randomly explored the specific residue substitutions in the endonuclease domain of the TAHV L segment.

1. Introduction

Ťahyňa virus (TAHV), member of California complex (CAL) of the genus Orthobunyavirus, family *Bunyaviridae*, was isolated by Bárdoš and Danielová from mosquitoes in Slovakia in 1958 (Bárdoš and Danielová, 1959). TAHV is closely related to the North American La Crosse (LACV) and the Snowshoe Hare viruses (Casals, 1962) and is a frequent cause of human infections (Valtice fever, VF), varying in clinical presentation and severity (Sluka, 1969); the most common form of VF is an inapparent or influenza-like disease. A pulmonary form of VF is manifested by atypical pneumonia, bronchopneumonia, or pleuritis. An anginose form includes symptoms of catarrhal, lacunary or follicular tonsillitis, and, if associated with another infection, an anginose form can be severe. Less frequently, the TAHV infection may be associated with rheumatic disease affecting the joints, abdominal pain of various intensity, or cause symptoms of meningitis serosa (Bárdoš and Sluka, 1963; Mittermayer et al., 1964; Sluka, 1969). In 1970–80s it was demonstrated that every seventh influenza-like disease and every fifth case of aseptic meningitis in children during summer months in the Czech Republic are caused by TAHV (Bárdoš et al., 1980). The highest TAHV seropositivity is found in central Europe (Czech Republic, Slovakia, Austria, and Hungary), but TAHV antibodies are found in humans in southern and northern Europe, in Africa, and in Asia (Bardos and Sefcovicova, 1961; Kunz, 1965; Lu et al., 2009; Gao et al., 2010; Jentes et al., 2010).

Like other bunyaviruses, TAHV is an enveloped spherical virus, approximately 100 nm in diameter with a tri-segmented, nega-

Abbreviations: AKAV, Akabane virus; BUNV, Bunyamwera virus; CAL, California complex; CHATV, Chatanga virus; DPBA, 2,4-dioxo-4-phenylbutanoic acid; d_S, synonymous nucleotide substitution rate; d_N, non-synonymous nucleotide substitution rate; ED, endonuclease domain; INKV, Inkoo virus; L, large genomic segment; LACV, La Crosse virus; M, medium genomic segment; ML, maximum likelihood; MP, maximum parsimony; N, nucleocapsid protein; NSs, nonstructural protein encoded by S segment; OROV, oropouche virus; ORF, open reading frame; OTU-like, ovarian tumour-like; PELE, protein energy landscape exploration; RMSD, root mean square deviation; S, small genomic segment; TAHV, Ťahyňa virus; ts, temperature-sensitive; VF, valtice fever.

* Corresponding author at: Department of Virology, Veterinary Research Institute, Hudcova 70, CZ-62100 Brno, Czech Republic. Tel.: +420 38 777 5451; fax: +420 38 531 0388.

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.09.023

tive-sense RNA genome of approximately 13 kb in length. The three segments are designated by their size, small (S), medium (M), and large (L). The S segment encodes two proteins in overlapping reading frames, namely, a nucleocapsid protein (N) and a non-structural protein (NSs). The M genomic segment encodes for one polyprotein that is co-translationally cleaved into three proteins: surface glycoproteins G_N and G_C, and a non-structural protein NSm. The L segment encodes for an RNA-dependent RNA polymerase (Elliott, 1990). TAHV is mainly transmitted by mosquitoes of the genus *Aedes*, specifically *Aedes vexans* and *Aedes cantans*. Hares, rabbits, hedgehogs, and rodents serve as amplifying hosts (Hubálek et al., 1993) for TAHV, and the virus can overwinter in mosquito eggs or in live mosquitoes (Danielová and Minář, 1969; Danielová and Ryba, 1979).

Although TAHV is a common and an important human pathogen, limited studies have invested to understand its genetics (Quinan et al., 2008; Kilian et al., 2010; Bennett et al., 2011). In our previous work, we characterized a genetic variability of biologically diverse TAHV strains in the S and M genomic segments (Kilian et al., 2010). In this study, we sequenced the L segments of 11 biologically diverse TAHV strains, completed their full-length genomic sequence, and discuss the relation of genetic variation to phenotype of TAHV. We also explore the all-atom energy landscape of a viral polymerase inhibitor using the crystal structure of the endonuclease domain of LACV (Reguera et al., 2010).

2. Materials and Methods

Eleven TAHV strains were analyzed in our study. These included 9 low-passage field TAHV isolates, one laboratory maintained strain 236, and one highly attenuated 181/57. For information on biological characteristics of the analyzed TAHV strains see Supplementary Table 1.

Viral RNA was isolated from mouse brain suspensions using Qiagen QIAamp Viral RNA Mini Kit (Qiagen, Germany) according to instructions of the manufacturer. Reverse transcription reaction was performed by RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas, Canada) according to recommendation in the manual with 1 μl of sense primer (0,1 mM) TL 3′F (primer sequences available upon request). PCRs were performed by Plain PPP Master Mix ((150 mM Tris–HCl, pH 8.8, 40 mM (NH4)SO4, 0.02% Tween 20, 5 mM MgCl2, 400 M dNTPs, 100 U/ml Taq DNA polymerase; TopBio, Czech Republic) with 2 μl of the prepared cDNA, 1 μl of sense and 1 μl of antisense primer (for primer sequences see Supplementary Table 2). The reaction conditions were following: first denaturation 5 min at 95 °C followed by 30 cycles: 95 °C/30 s, 50 °C/30 s, 72 °C/1 min, after 30 cycles was the final amplification at 72 °C for 10 min. Purified fragments of DNA were directly sequenced by ABI Prism 3130 xL (Applied Biosystems) sequencer. Sequences were aligned and assembled using Sequencher™ 4.9 built 4759 software. All derived sequences were deposited in GenBank database under the accession numbers KF361874–KF361884.

Alignment of the complete TAHV L segment sequences with other members of the genus Orthobunyavirus (La Crosse virus – LACV, Snowshoe hare virus – SSHV, Chatanga virus – CHATV, Bunyamwera virus – BUNV, Oropouche virus – OROV, and Akabane virus – AKAV) family was performed using the ClustalW multiple alignment implemented in BioEdit 7.0.9.0 sequence alignment editor. Basic bioinformatics data were acquired with MEGA4 software (Zhang and Gu, 1998; Tamura et al., 2007) and percentual differences among European and Asian strains were acquired using Geneious® Pro 5.6.5 software.

To analyze the nucleotide substitution rate (synonymous (d_S) and the non-synonymous (d_N) substitutions per site) for the 11 TAHV strains, we first performed a codon-based alignment for both datasets using the GUIDANCE server (Penn et al., 2010). We then submitted each codon alignment to the SNAP server (Korber, 2000) that uses the Nei and Gojobori method (Nei and Gojobori, 1986). We then used the ratio d_N/d_S to identify whether a positive (>1) or purifying selection (<1) occured during it evolution.

We submitted the nucleotide sequences for strains T92B, T94, T4020 and T1014 to the NCBI Open Reading Frame Finder (ORF) online server (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to verify and edit the sequences to an appropriate start-stop codon. We were interested in the endonuclease domain of the L-segment, so we used the translated amino acid sequences provided by the ORF Finder for this portion. The endonuclease portion of the L-segment for strains T92B, T94, T4020 and T1014 were submitted to the Swiss-Model server (Arnold et al., 2006; Kiefer et al., 2009) that provided us with adequate homology models. The models were then refined and prepared using the Schrodinger's Protein Wizard (Li et al., 2007), which optimizes the entire hydrogen bond network.

To investigate the ligand migration of an antiviral inhibitor and the LACV crystal structure we used the Protein Energy Landscape Exploration (PELE) server (Madadkar-Sobhani and Guallar, 2013). The server provides a ready-made script for an unconstrained ligand exploration and binding site search. The only parameters that we altered in the script was to constrain the native pose of Mn^{2+}, the residues forming polar contacts, and two water molecules. PELE incorporates large perturbations followed by a relaxation of the protein–ligand conformation. Next, the Monte Carlo Metropolis criterion is used to either accept or reject the conformational changes (Borrelli et al., 2005).

Phylogenetic analyses were performed using amino acid sequences that were aligned in MAFFT version 6 (Katoh et al., 2002, 2005) using E-INS-I algorithm. Unaligned and ambiguously aligned positions were deleted from datasets. Phylogenies were inferred using maximum likelihood (ML), maximum parsimony (MP) and Bayesian inference (BI), respectively in the softwares PhyML 3.0 (Guindon and Gascuel, 2003), PAUP* 4.0 (Swofford, 2000) and MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Phylogenetic model (ML = LG + Γ(6 categories)+F; BI = WAG + Γ(6 categories) was estimated using ProtTest 2.4 (Abascal et al., 2005). Plot of nucleotide similarity with the strain 236 was performed using Simplot 3.5.1 software (http://sray.med.som.jhmi.edu/SCRoftware/simplot/) using kimura 2-parameter model, window size 400 and step size 10 b.

3. Results and discussion

In this study we analyzed the complete sequence of the L genomic segment of TAHV CAL group in the genus Orthobunyavirus within the *Bunyaviridae* family. Among our set of isolates, we observed the same length of the L segment as described previously; i.e., 6976 bases (Bennett et al., 2011). We used two datasets to analyze nucleotide substitution rate (synonymous or non-synonymous). Dataset 1 included the 11 TAHV L segment sequences from our study, other TAHV strains, plus related bunyaviruses (LACV, INKV, SSHV, and CHATV) to detect the substitution rate among TAHV strains. Dataset 2 included the 11 TAHV L segment strains from our study together with 24 different strains of LACV and four strains of Chatanga virus. Because TAHVs belong to a large family of viruses, Dataset 2 was designed to analyze the substitution rate between TAHV strains with LACVs and among related viruses. We noted that there was a higher synonymous nucleotide substitution rate (d_S) compared with non-synonymous (d_N) among the two datasets. (The specific non-synonymous substitutions throughout the L segment will be discussed later in further detail.) The ratio d_N/d_S reveals that both datasets have undergone a purify-