Ticks and Tick-borne Diseases 6 (2015) 38–46

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

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

Original article

Full genome sequences and molecular characterization of tick-borne encephalitis virus strains isolated from human patients

Petra Formanová[a,b], Jiří Černý[c,d], Barbora Černá Bolfíková[e], James J. Valdés[c], Irina Kozlova[f,g], Yuri Dzhioev[f,g], Daniel Růžek[a,b,c,d,*]

[a] Department of Virology, Veterinary Research Institute, Hudcova 70, CZ-62100 Brno, Czech Republic
[b] Faculty of Science, Masaryk University, Kotlářská 267/2, CZ-61137 Brno, Czech Republic
[c] Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
[d] Faculty of Science, University of South Bohemia, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
[e] Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Kamýcká 126, CZ-16521 Prague, Czech Republic
[f] Institute of Biomedical Technology, Irkutsk State Medical University of Russian Ministry of Health, Krasnogo Vosstanija 1, Irkutsk 664003, Russia
[g] FSSFE Scientific Centre of Family Health and Human Reproduction Problems, Siberian Branch of the Russian Academy of Medical Sciences, Timirjazeva Street 16, 664003 Irkutsk, Russia

ARTICLE INFO

Article history:
Received 17 June 2014
Received in revised form 5 September 2014
Accepted 5 September 2014
Available online 11 October 2014

Keywords:
Tick-borne encephalitis virus
Tick-borne encephalitis
Genome analysis
Human patients

ABSTRACT

Tick-borne encephalitis virus (TBEV) causes tick-borne encephalitis (TBE), one of the most important human neuroinfections across Eurasia. Up to date, only three full genome sequences of human European TBEV isolates are available, mostly due to difficulties with isolation of the virus from human patients. Here we present full genome characterization of an additional five low-passage TBEV strains isolated from human patients with severe forms of TBE. These strains were isolated in 1953 within Central Bohemia in the former Czechoslovakia, and belong to the historically oldest human TBEV isolates in Europe. We demonstrate here that all analyzed isolates are distantly phylogenetically related, indicating that the emergence of TBE in Central Europe was not caused by one predominant strain, but rather a pool of distantly related TBEV strains. Nucleotide identity between individual sequenced TBEV strains ranged from 97.5% to 99.6% and all strains shared large deletions in the 3′ non-coding region, which has been recently suggested to be an important determinant of virulence. The number of unique amino acid substitutions varied from 3 to 9 in individual isolates, but no characteristic amino acid substitution typical exclusively for all human TBEV isolates was identified when compared to the isolates from ticks. We did, however, correlate that the exploration of the TBEV envelope glycoprotein by specific antibodies were in close proximity to these unique amino acid substitutions. Taken together, we report here the largest number of patient-derived European TBEV full genome sequences to date and provide a platform for further studies on evolution of TBEV since the first emergence of human TBE in Europe.

Introduction

Tick-borne encephalitis (TBE) is the most important arboviral infection in Europe and Central and Eastern Asia. More than 13,000 human TBE cases are reported annually (Mansfield et al., 2009). The disease is caused by tick-borne encephalitis virus (TBEV), a member of the genus *Flavivirus*, family *Flaviviridae* (Mansfield et al., 2009).

TBEV is an enveloped virus with approximately 11 kb long single-stranded RNA genome of positive polarity. The genomic RNA contains one open reading frame (ORF) encoding single polyprotein. It is co- and post-translationally cleaved by viral and host proteases into three structural (capsid (C), membrane (M; derived from its precursor, prM) and envelope (E)) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) (Mansfield et al., 2009). Structural proteins are responsible for packaging of virus genome and budding of viral capsids through cellular membranes. Non-structural proteins catalyze replication of viral genome and regulate host-antiviral response.

The main ORF is flanked with 5′ and 3′ non-coding regions (NCRs). The 5′ NCR has a length of approximately 100 bp and is relatively homogenous on both size and sequence. The 3′ NCR is

* Corresponding author at: Department of Virology, Veterinary Research Institute, Hudcova 70, CZ-62100 Brno, Czech Republic. Tel.: +420 5 3333 1101; fax: +420 5 4121 1229.
E-mail address: ruzekd@paru.cas.cz (D. Růžek).

http://dx.doi.org/10.1016/j.ttbdis.2014.09.002
1877-959X/

extremely heterogeneous in length (751 bp in TBEV strain Neudoerfl, 445nt in TBEV strain Hypr) (Wallner et al., 1996). Rarely, the 3′ NCR of some TBEV strains contains a shorter poly(A) tail (Wallner et al., 1996; Frey et al., 2014). Both NCRs contain conserved secondary structures that are supposed to be involved in TBEV genome amplification, translation and packaging (Gritsun et al., 1997).

Based on phylogenetic analysis, TBEV can be divided onto three subtypes: the European subtype (Eu-TBEV), the Siberian subtype (S-TBEV), and the Far Eastern subtype (FE-TBEV) (Ecker et al., 1999). Members of these three subtypes differ in their geographical distribution, virulence, and clinical severity of caused diseases (Mansfield et al., 2009).

Although the medical and economic impact of TBE is high, the TBEV strains isolated from patients remain largely unstudied and only a few complete genome sequences of human Eu-TBEV strains have been reported until now. This paucity is caused by the difficulty in obtaining TBEV isolates from humans – the virus can be isolated from blood during the first (nonspecific) phase of the infection or from *post mortem* brain tissue. During the neurological phase of the infection, the virus is rarely present in the blood or the cerebrospinal fluid of the patients (Růžek et al., 2010) and most isolation attempts are usually unsuccessful.

Almost all Eu-TBEV strains with known genome sequence were isolated from ticks or rodents. However, analysis of complete nucleotide sequences of strains isolated from patients with variable disease severities is crucial for detection of mutations in the TBEV genome that determine the pathogenicity for humans (Belikov et al., 2014). Currently, only three complete Eu-TBEV genome sequences are available, which were isolated from human patients. Strain “Hypr” was isolated in 1953 from the blood of a diseased young boy in Czechoslovakia (Pospíšil et al., 1954; Wallner et al., 1996). Strain “Est3476” was obtained from a serum sample of patient from Estonia (Golovljova et al., 2004). Finally, strain “Ljubljana 1” was isolated in 1992 from blood of a TBE patient from Slovenia (Fajs et al., 2012). The largest set of European patient-derived TBEV sequences was provided by analysis of E gene sequences of 15 strains and NS5 gene sequences of 17 strains (Fajs et al., 2012). Recently, a comparison of 34 genomes of FE-TBEV strains isolated from patients with different disease severities identified specific mutations responsible for differences in pathogenicity of FE-TBEV strains (Leonova et al., 2014; Belikov et al., 2014). However, there are large differences in sequence of FE-TBEV and Eu-TBEV that also underlines a need of analysing patient-derived Eu-TBEV complete genomes.

The TBEV of Central Europe was first isolated in 1948 in the former Czechoslovakia (Krejčí, 1949; Gallia et al., 1949). The TBEV strains analyzed in this study belong, therefore, together with other strains from the late 1940s and early 1950s, are the oldest human TBEV isolates in Europe. Here, we report a total of five full genome sequences from patient-derived European TBEV strains to date. We also provide a platform to further analyze TBEV evolution and its antigenic properties since the first TBEV emergence in Europe.

Materials and methods

Five archival low-passage TBEV strains were selected for the full genome sequence analysis. These strains were isolated from the blood of patients hospitalized with TBEV infection during the TBEV outbreak in 1953 in Central Bohemia (Czechoslovakia). All patients had severe course of the TBE.

RNA was isolated from 20% suckling mouse brain suspension using QIAamp Viral RNA Mini Kit (Qiagen). Reverse transcription was performed using ProtoScript® First Strand cDNA Synthesis Kit (New England Biolabs). The 35 overlapping DNA fragments were produced by PPP Master Mix (Top-Bio, sequence of primers is available on request) as described previously (Růzek et al., 2008). The PCR products were then sequenced directly by commercial service (SEQme, Czech Republic). The deduced whole genome sequences were deposited in the GenBank database under accession numbers: KJ922512–KJ922516. Both nucleotide and deduced amino acid sequences were analyzed using BioEdit Sequence Alignment Editor, version 7.2.0 (Hall, 1999) and MultAlin (Corpet, 1988), aligned by Muscle in MEGA version 5 (Tamura et al., 2007). For complete sequence comparisons we used 60 complete genomes of TBEV together with Turkish sheep encephalitis virus (TSEV; GenBank accession number: DQ235151.1), Spanish sheep encephalitis virus (SSEV; DQ235152.1) and Louping ill virus (LIV; Y07863.1) deposited in GenBank database. For detection of selection pressure acting on individual genes we calculated the ratios of non-synonymous and synonymous nucleotide substitutions per site (dN/dS) of the available TBEV sequences using MEGA version 5 (Tamura et al., 2007).

The predicted secondary structure of the NCRs was produced using Mfold server (http://mfold.rna.albany.edu) under default conditions.

Best fitting model of nucleotide substitutions was tested in jModelTest (Darriba et al., 2012). The general time reversible (GTR) model was selected as the best fitting model. Bayesian phylogenetic analysis was performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). Bayesian analysis consisted of two runs with four chains (one cold and three heated), and was run for 10 million generations sampled every 500 generations. The first 25% of samples were discarded as a burning period. The average standard deviation of split frequencies was 0.001 showing convergence of all chains.

We used 1SVB to depict structure of TBEV protein E (Rey et al., 1995). Structures of proteins NS1, NS3, and NS5 were modeled by homology modeling on Phyre2 server (Kelley and Sternberg, 2009) and proteins were modeled according 4O6C (Akey et al., 2014), 2VBC (Luo et al., 2008), and 4K6M (Lu and Gong, 2013) templates. Molecular rendering was done using PDB Swiss Viewer (Guex and Peitsch, 1997). The TBEV protein E crystal structure and predicted models were prepared and refined by adding hydrogen atoms, optimization of the hydrogen-bond network, followed by a full minimization of the system to remove steric clashes (i.e., overlapping atoms) using the Schrodinger's Maestro software (Li et al., 2007). The prepared structures were then submitted to the ElliPro server to predict epitope(s) position(s). The ElliPro server uses the tertiary structures to predict epitope regions based on their particular scoring function (Ponomarenko et al., 2008). For the antibody–antigen docking we used the SwarmDock server (Torchala et al., 2013a,b; Torchala and Bates, 2014) that incorporates flexible protein–protein docking by exploring around the Cartesian center of mass of the receptor (the antigen) and including minimization steps for the whole system. Once energy favorable poses are generated they minimized again sent to the user.

Results and discussion

We have sequenced and analyzed the complete genomes of 5 Eu-TBEV strains Skrivanek, Petracova, Vlasaty, Tobrman and Kubinova isolated from patients with severe TBE in 1953. Nucleotide identity between individual sequenced TBEV strains ranged 97.5–99.6%. All isolates, therefore, represent unique strains, although they were isolated during the same season and in the same geographic region. The length of the nucleotide sequence of the genomes ranged from 10,777 to 10,979 nucleotides. The differences in genome length were due to the variable length of the 3′ NCR. The ORF of all isolates were of standard length (10,245 nt).