

The use of next-generation sequencing for genetic characterization of tick-borne encephalitis virus variants from Western Siberia

Tkachev S.E.^{1,2*}, Shaikhutdinov N.M.^{1,3}, Shigapova L.Kh.¹, Chicherina G.S.⁴, Yakimenko V.V.⁵, Shagimardanova E.I.¹

¹ Research Center "Regulatory Genomics" of Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia

² Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia

³ Skolkovo Institute of Science and Technology, Moscow, Russia

⁴ Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia

⁵ Omsk Research Institute of Natural Focal Infections, Omsk, Russia

* sergey.e.tkachev@gmail.com

Key words: tick-borne encephalitis virus, Western Siberia, subtypes, lineages

Introduction: Tick-borne encephalitis virus (TBEV), a member of the *Flaviviridae* family, is the causative agent of a severe disease of the human central nervous system. Currently, in accordance with classification TBEV is divided into three subtypes: Far Eastern, Siberian (TBEV-Sib) and European [1]. Until recently, three genetic lineages were described for TBEV-Sib: Zausaev, Vasilchenko, and Baltic; subsequent studies showed that there are at least two more lineages, Obskaya and Bosnia [2, 3].

In the majority of TBEV collections genetic studies, sequencing of only genome fragments is used, which limits the use of the data obtained to assess genetic diversity, the distribution of certain virus variants, phylogeography, as well as the patterns that determine the evolution of their genomes, and only complete genome sequences provide the necessary information. The use of "classical" approaches based on Sanger sequencing for this is inefficient and time consuming. The solution to this problem could be the use of high-throughput sequencing (next-generation sequencing, NGS) to analyze large sets of TBEV strains samples from virus collections.

Currently, only about 250 TBEV complete genome sequences are presented in the GenBank databases (<https://www.ncbi.nlm.nih.gov/nucleotide/>) (and only about 30 of them for TBEV-Sib), so the task of replenishing data on the genomes of various TBEV genetic variants is a crucial task.

Aim of the study: The aim of this work was to develop approaches to high-throughput sequencing of TBEV strains, obtain TBEV complete genome sequences of strains from different regions of Western Siberia, and perform molecular genetic analysis of the obtained data in comparison with the available TBEV sequences from the GenBank database.

Materials and Methods: For the study, 85 TBEV strains previously isolated by bioassay method from various natural samples (different ticks and small mammals species) in various regions of Western Siberia (Novosibirsk oblast, Omsk oblast, Khanty-Mansiysk Autonomous District, the Altai Republic of the Russian Federation, as well as Kazakhstan) were taken. From the isolated genomic RNAs, genomic libraries were constructed and sequenced using Miseq (Illumina). After assessing the quality of the

data and trimming, genomes were assembled either *de novo* (with sufficiently high coverage) or using the data mapping method with published reference TBEV genomes (with low coverage). For further analysis, TBEV complete genome sequences from the GenBank database corresponding to various subtypes and genetic lineages of the virus were used.

Results: Analysis of determined complete genome sequences of TBEV strains isolated in various regions of Western Siberia revealed predominantly TBEV-Sib strains, with single findings of TBEV of the Far Eastern subtype in the Altai Republic and the Novosibirsk Region, and of the European subtype in the Omsk Region. Zausaev lineage was found to be the most represented among the studied strains, and the Vasilchenko lineage was represented to a lesser extent, which correlates with our earlier data. Also, strains of the Baltic lineage were discovered and characterized in the Novosibirsk and Omsk oblasts, as well as, for the first time, in Kazakhstan. No representatives of the Obskaya lineage previously described in the indicated territory were found, which, as previously shown, is most genetically different from other TBEV-Sib lineages. Analysis of the chronogram demonstrated that within the TBEV-Sib, the Obskaya lineage diverged from the common ancestor the earliest, after that the Bosnia lineage was separated, then the Baltic lineage, and the Zausaev and Vasilchenko lineages diverged most recently.

Acknowledgements: The study was part of Kazan Federal University Strategic Academic Leadership Program (PRIORITY-2030) and funded by the Pfizer Quality Improvement project No. 65238411 "Optimization of methods for mass complete genome sequencing of tick-borne encephalitis virus strains".

References

1. Virus taxonomy: classification and nomenclature of viruses. In: King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J. (Eds.). Ninth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier Academic Press, 2012.
2. Tkachev S.E. et al. New genetic lineage within the Siberian subtype of tick-borne encephalitis virus found in Western Siberia, Russia. *Infect Genet Evol.* 2017;56:36-43. doi: 10.1016/j.meegid.2017.10.020.
3. Tkachev S.E. et al. Genetic diversity and geographical distribution of the Siberian subtype of the tick-borne encephalitis virus. *Ticks Tick Borne Dis.* 2020;11(2):101327. doi: 10.1016/j.ttbdis.2019.101327.4.