DOI 10.18699/SBB-2022-506

Genetic traits of the delta variant in the Russian Federation and its spread during the epidemic caused by the SARS-CoV-2 virus

Ayginin A.A.[1]*, Vinogradov K.S.[1], Bulusheva I.A.[2], Matsvay A.D.[1], Mikhaylov I.M.[1], Luparev A.R.[1], Polyakova V.A.[1], Shipulin G.A.[1]

[1] *Centre for Strategic Planning and Management of Biomedical Health Risks of the Federal Medical Biological Agency, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*

** AAyginin@cspmz.ru*

Key words: COVID-19, SARS-CoV-2, delta, mutations, G1048T, C27527T

Motivation and Aim: The COVID-19 epidemic caused by the SARS-CoV-2 virus has affected most countries globally and is widely studied by scientists worldwide. The first cases of infection with this virus were identified in December 2019 in China in Wuhan Province [1]. Over the past two years, the virus has repeatedly mutated, producing many variants that have spread worldwide. Evolution of the delta variant of coronavirus may occur independently in different countries and have specific characteristics. This study describes the characteristic genetic variants of the delta variant of coronavirus in Russia.
Methods and Algorithms: To analyze the genetic variants of the delta variant of coronavirus, both samples from the GISAID database (database version dated July 31, 2021) and samples collected by the Centre for Strategic Planning of FMBA laboratory in various regions of Russia were used. For the samples sequenced by the Centre for Strategic Planning of FMBA laboratory, the virus genomes were assembled in the following stages: trimming of adapters and low-quality bases; primers trimming; mapping of reads to the MN908947 reference genome; read filtering; SNV/indel calling; consensus building. For samples obtained from the GISAID database, filtration was conducted according to the following algorithm: sequences were aligned to the reference genome, the first and last 100 nucleotides were removed, sequences were filtered by length (at least 29,000 bp) and the number of uncertain bases (not exceeding 3000 N). The resulting genomes were clustered using the cd-hit [2] program, which reduced the database size for further analysis. Phylogenetic tree was constructed using the RAxML software [3].
Results: Russian isolates of the delta variant of the virus were found in 4 different branches of the phylogenetic tree. At the same time, most (95 %) of the Russian samples of the delta variant formed a separate branch, defined by G1048T and C27527T substitutions. Unlike other countries, where the proportion of G1048T mutation among delta strain samples did not exceed 7 % on average at the beginning of the spread of the delta variant, the corresponding value in Russia increased up to 100 % in early May 2021. The same result was observed for the C27527T substitution. By the end of July 2021, the proportion of G1048T+C27527T substitution in the Russian samples was 93 % while this number in the worldwide samples did not exceed 1 %. One of the countries in which these two mutations were also observed shortly before Russia could become a source of import of the delta strain to the country. Assuming that the candidates for the import of the variant to Russia are Italy, Germany, Israel, Singapore, Japan, and Switzerland.

Conclusion: The first double mutant sample found in Russia was obtained on April 19, 2021. The tree constructed using the RAxML program suggests that the G1048T mutation appeared first, after which the G27527T substitution appeared inside this cluster. However, this conclusion requires an in-depth study to confirm. These mutations may affect coronavirus transmissivity. The G1048T mutation (NSP2:K81N) is located in the NSP2 protein, and its function has not been properly studied to date. The protein participates in viral replication and forms a protease complex required to cleave the ORF1ab polyprotein [4]. Notably, the NSP2 protein is considered a potential target for COVID-19 medical therapy; therefore, mutations arising in it may affect the effectiveness of the medicines being developed. The C27527T mutation (ORF7A:P45L) is located in the ORF7a protein. This protein interacts with the immune cells and may affect the strength of the immune response. Recently, Zhou et al. proved that ORF7a SARS-CoV-2 binds to CD14+ monocytes [5]. The data obtained in this study prove that the ORF7a protein is a modulating factor for immune cells binding to the virus particles and an inflammatory response, indicating the possibility of its use as a therapeutic target for the COVID-19 treatment. Based on the analysis of emerging mutations, tracing the evolution of this variant in Russia is possible. Additionally, data on existing mutations and their combinations may be used to determine the sources of the spread of coronavirus in different countries.

References

1. Wu F. et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. DOI 10.1038/s41586-020-2008-3.
2. Li W., Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659. doi: 10.1093/bioinformatics/btl158.
3. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinforma Oxf Engl.* 2014;30(9):1312-1313. doi: 10.1093/bioinformatics/btu033.
4. Raj R. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochem Biophys Rep.* 2021;25:100847. doi: 10.1016/j.bbrep.2020.100847.
5. Zhou Z. et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience*. 2021;24(3):102187. DOI 10.1016/j.isci.2021.102187.