Phylogenetic Changes in SARS-CoV-2 Virus in Bosnian-Herzegovinian Population Over the Period of Two Years

Naida Mulahuseinovic1, Lana Salihefendic1, Dino Pecar1, Ivana Ceko1,3, Selma Durgut1, Enis Kandic1, Vesna Hadziavdic2, Rijad Konjhodzic1,3
1ALEA Genetic Center, Sarajevo, Sarajevo, Bosnia and Herzegovina
2Department of Biology, Faculty of Natural Sciences and Mathematics, University of Tuzla, Tuzla, Bosnia and Herzegovina
3Department of Genetics and Bioengineering, International Burch University Sarajevo, Ilidza, Bosnia and Herzegovina

Corresponding author: Naida Mulahuseinovic, ALEA Genetic Center, Olovska 67, 71000 Sarajevo, Bosnia and Herzegovina Mob.: +38761587070. E-mail: naida.mulahuseinovic@agc.ba. ORCID ID: http://www.orcid.org/0000-0001-8776-9143.

ACTA INFORM MED. 2023 MAR 31(1): 57-61
Received: Dec 01, 2022
Accepted: Feb 04, 2023

ABSTRACT

Background: All viral genomes, including the SARS-CoV-2 virus, mutate over time, and some of these mutations can affect the characteristics of the virus, such as the ease of spread, the severity of the patient’s clinical picture, or the effect of vaccines, therapeutic drugs, diagnostic tools or other measures of public health and social protection. Because of all the above, it is imperative to carry out continuous sequencing of this pathogen.

Objective: The main goal of this research was to obtain the highest quality genomic sequences of the SARS-CoV-2 virus, to compare the obtained sequences with the reference Wuhan-Hu-1 sequence and to obtain a high-quality genomic alignment in order to reconstruct the appropriate phylogenetic tree.

Methods: For the purposes of this research, a next-generation semiconductor sequencing method was chosen. In this research, a total of 47 samples of nasopharyngeal and oropharyngeal swabs from patients from the human population of Bosnia and Herzegovina with a clinical diagnosis of COVID-19 were collected.

Results: In the processed 47 samples, there are several monophyletic groups on the constructed phylogenetic tree, of which one sample belongs to the same monophyletic group as the Wuhan-Hu-1 reference sequence.

Conclusion: The greater number of samples is needed for a more comprehensive approach. Therefore, the results of this research can act as a guideline for the design of effective measures and strategies in order to solve problems regarding future pandemics as efficiently as possible.

Keywords: viral genome sequencing, lineage shift, phylogeny, SARS-CoV-2, next generation sequencing.

1. BACKGROUND

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in Wuhan, Hubei Province, China, in December 2019 (1) and spread rapidly throughout the world infecting millions of people to this day (2). Rapid methods of sequencing to monitor genetic changes were essential for understanding the evolution and dynamics of virus transmission (3). Rambaut et al. (2020) have designed a nomenclature for SARS-CoV-2 lineages, and in that way it was possible to detect changes and similarities between different outbreaks of the epidemic. The nomenclature was named Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin) (4). According to this nomenclature, two main lineages can be distinguished: lineages A and B. Lineage A represents the beginning of the pandemic in December 2019 and lineage B that contains Chinese sequence at its core with different branching as the pandemic grew and spread. Lineage B.1 dates from January 2020 and is connected to the pandemic outbreak in Italy; lineage B.1.17 dates from September, 2020; lineage B.1.351 emerged in December, 2020, B.1.1.28.1 lineage was first detected in December 2020 and B.1.617 that emerged in March, 2021(5-8). Pangolin lineage B.1.1.529 and its descendants BA.2, BA.1, BA.4, BA.3 and BA.5 firstly emerged in Botswana in November 2021 and this variant soon replaced the dominant Indian variant all around the world (9). Information obtained from constant sequencing and determination of new SARS-CoV-2 lineages can help and improve reduction of pandemic effects, as well as have an impact on vaccination...
strategies, therapies in COVID-19 patients, etc (10).

First case of infection with SARS-CoV-2 virus in Bosnia and Herzegovina was detected on March 5th, 2020 in Banja Luka. SARS-CoV-2 virus has been sequenced for the first time in June 2020 in ALEA Genetic Center (Sarajevo, B&H). ALEA Genetic Center has hitherto been sequencing and determining SARS-CoV-2 changes regularly. Goletić et al. (2021) have published first results regarding the phylogenetic pattern of SARS-CoV-2 in Bosnia and Herzegovina (11).

2. OBJECTIVE

The aim of this study is the implementation of the next-generation semiconductor sequencing technology to obtain the highest possible coverage of the SARS-CoV-2 genomic sequence on samples of nasopharyngeal and oropharyngeal swabs of the Bosnian-Herzegovinian human population with a clinical diagnosis of COVID-19, and to determine which lineages of this virus were present in the human population of Bosnia and Herzegovina in the period of research.

3. MATERIALS AND METHODS

For the purposes of this research, a total of 47 samples of nasopharyngeal and oropharyngeal swabs immersed in viral transport medium were taken from Bosnian-Herzegovinian population with a clinical diagnosis of COVID-19 in the period from May 2020 to March 2022. All laboratory work from RNA extraction to sequencing of SARS-CoV-2 virus was performed in ALEA Genetic Center, Sarajevo. RNA extraction was performed using an automatic nucleic acid extractor based on magnetic particle adsorption separation technology—Tianlong Technology GeneRotex 96 (Tianlong Science & Technology, Shaanxi, China) according to manufacturers instructions. Detection of SARS-CoV-2 in samples was performed using the LabGun™ COVID-19 ExoFast RT-PCR Kit according to manufacturers instructions (LabGenomics Co., Ltd, Bruxelles, Belgium) on Bio-Rad Cfx96 Real-Time PCR Detection System instrument (Bio-Rad Laboratories, California, USA). Selection of samples for sequencing was based on their Ct value obtained by real-time PCR method. Positive samples that had a Ct value less than 24 were selected for sequencing. cDNA synthesis was performed using the Invitrogen™ SuperScript™ Vilo™ cDNA Synthesis Kit according to manufacturers instructions (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Ion AmpliSeq™ Library Kit Plus (Thermo Fisher Scientific, Massachusetts, USA) with panel Ion Ampliseq SARS-CoV-2 Research Panel (Thermo Fisher Scientific, Massachusetts, USA) was used for NGS library preparation according to manufacturers instructions. It was determined that the optimal number of cycles in the annealing phase of target cDNA amplification during library preparation is 37. Quantification of libraries was done by qPCR method on QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems, Massachusetts, USA) using Ion Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific, Massachusetts, USA). Sequencing was performed using the Ion GeneStudio™ S5 System (Thermo Fisher Scientific,
Phylogenetic Changes in SARS-CoV-2 Virus in Bosnian-Herzegovinian Population Over the Period of Two Years

Massachusetts, USA). Sequencing using Ion Torrent S5 platform (Thermo Fisher Scientific, Waltham, Massachusetts, USA) enables multi parallel sequencing with deep coverage and it is suitable for high throughput testing (12).

Nextstrain software was used to confirm the lineages and clades of the SARS-CoV-2 virus that were present in Bosnia and Herzegovina at the time of the research. A phylogenetic tree was constructed using MEGA 11 software. This phylogenetic tree represents a visual representation of the assumed evolutionary relationship between the 48 analyzed sequences (47 sequences processed in this research and an additional reference sequence).

4. RESULTS

For the construction of the phylogenetic tree, the Maximum-likelihood (ML) method was used, according to the Jukes-Cantor substitution model. Nearest-Neighbor-Interchange (NNI) was chosen as the tree inference model. On the phylogenetic tree, sample numbers are marked with following colors: green numbers signify the samples that were collected in 2020, blue numbers signify the samples that were collected in 2021 and red numbers signify the samples that were collected in 2022 (Graph 1).

On the constructed phylogenetic tree, the phylogeny of the 47 sequences obtained in this study was interpreted in relation to the Wuhan-Hu-1 reference sequence, and for this reason the root of the tree was set according to the reference sequence. Based on this, it can be concluded that one sample belongs to the same monophyletic group as the reference sequence. This sample, which was collected on January, 2021, belongs to the 20A clade, and is the only sample belonging to the B.1.146 lineage.

In 2020, four samples were sequenced, out of which two samples belong to lineage B.1 and they were collected in April and May of the same year, while the other two belong to lineages B.1.258 and C.35 which were collected in November 2020.

In January 2021, B.1.146, B.1.36, B.1, B.1.258.17 and B.1.1.7 lineages have emerged in Bosnian-Herzegovinian population. Most of the detected SARS-CoV-2 lineages in February, 2021 belong to B.1.1.7 lineage, with exception of two samples that belong to B.1 lineage and one sample belonging to B.1.351 lineage (Beta variant). Our data suggest that no new lineages appeared in May 2021 in Bosnian-Herzegovinian population and that all samples sequenced in this period belong to B.1.1.7 lineage. New lineage emerged in June 2021 in the form of Delta (AY.122) and Gamma (P.1 and P.1.10) variants. It can be observed that all sequenced samples that are collected in July 2021 belong to Delta variant with different lineages; B.1.617.2, Ay.43, Ay.9.2, Ay.45 and Ay.34. Since the beginning of 2022, only the Omicron variant has been sequenced in this study with different lineages BA.1.1, BA.1.1.17.2 and BA.2.

5. DISCUSSION

One of the first mutations observed in SARS CoV 2 genome which are presumed beneficial for the virus survival was the D614G amino acid substitution in the S protein that provided a measurable transfer advantage. This was the defining amino acid substitution of the B.1 Pangolin lineage (13). Lineage B.1.258 has been circulating in Central Europe since August 2020 (14).

Lineage B.1 remained in Bosnia and Herzegovina at least until February 2021 when the last two samples of this lineage were sequenced according to this study, however this doesn’t mean that it did not persist in Bosnian-Herzegovinian population even longer. These two samples are in the same monophyletic group as samples that belong to B.1.1.7 lineage.

Highly mutated B.1.1.7 lineage was first identified in the United Kingdom and showed an unexpected step change in the evolution of the SARS-CoV-2 virus (15).

Most of the samples in this study that were sequenced in 2021, more precisely 14 of them belong to the B.1.1.7 lineage. The first sample of this lineage was collected in February 2021 and the last of them was collected in June, 2021. All of these samples are clearly separated in the one monophyletic group on the phylogenetic tree together with three other samples and two of them belong to already mentioned B.1 lineage while one of them belongs to B.1.351.

B.1.351 was identified in October 2020 and after that quickly became predominant lineage in South Africa. This lineage has ten changes (deletions and substitutions) in the spike protein (16). Only one sequenced sample in this study belongs to B.1.351 lineage and sampling was done in February 2021.

B.1.1.7 lineage quickly gave rise to another variant which led to a new global wave of the SARS-CoV-2 virus. However, instead of accumulating further antigenic change, the B.1.1.7 lineage was replaced by another, even more transmissible lineage, the B.1.617.2, which the World Health Organization named the Delta variant. Delta variant emerged in India in late 2020 and then caused a third wave of the pandemic in 2021 (17). The Delta variant has progressed effectively, expanding its ancestry into various subgroups or sublineages with AY prefix, such as: AY.1, AY.2, AY.3, AY.33, AY.34 (18).

In this study 5 sublineages of Delta variant in June and July, 2021 were found: AY.122, AY.43, AY.9.2, AY.45 and AY.34, additionally three samples of original B.1.617.2 lineage were sequenced in the same period. All of these samples are separated in one monophyletic group on phylogenetic tree along with sample that belongs to C.36.3 lineage or 20D clade. C.36.3 is also characterized by D614G spike mutation (19).

The Omicron variant is characterized by a set of previously observed and new mutations that provide both transmission and antigenic shift, which caused the fourth global wave of the SARS-CoV-2 virus. The omicron wave was initiated in most countries simultaneously by lineages BA.1 and BA.1.1 with the addition of the S protein mutation R346K, which are now being replaced by even more transmissible lineages, such as BA.2 (B.1.1.529.2) and BA.3 (B.1.1.529.3) (20). The Omicron variant is a representative of variants with hypermutated S protein and has adapted to some significant evolutionary changes (21). A possible source of such highly mutated viruses is at least one, and possibly more, infection of immunocompromised individuals or infection through animals (21, 22).

In this study all samples that were sequenced in 2022 belong to BA.1.1, BA.1.1.17.2 and BA.2 lineages and all of these samples are located on the same monophyletic group.

CONCLUSIONS
Phylogenetic Changes in SARS-CoV-2 Virus in Bosnian-Herzegovinian Population Over the Period of Two Years

From reported results, it can be concluded that the population of Bosnia and Herzegovina followed the world trend of the appearance and spread of global pandemic waves in a two-year period. Certain lineages persisted longer than others, especially at the beginning of the pandemic when the mutation rate of the SARS-CoV-2 virus was thought to be lower. On the other hand, some clades, such as clade 20D, remained in the Bosnian-Herzegovinian population for over a year with its sublineage C.36.3 which also took a place on the phylogenetic tree that is closer to samples from the period when it was sequenced than to samples that also belong to clade 20D.

The greater number of samples is needed for a more comprehensive approach. However, the results of this research can act as a guideline and a starting point for the design of effective measures and strategies to prevent the spread of the virus among the Bosnian-Herzegovinian population, as well as for the reconstruction of the appropriate phylogenetic tree in subsequent research.

This research opened up the possibility of establishing different platforms and adequate strategies based on NGS, which can be successfully used to trace the origin and understand the evolution of virus, research the chains of spread and transmission of epidemics, as well as to facilitate the development of effective and rapid molecular diagnostic tests and contribution to the search for clinical treatment and vaccine development.

And finally, Bosnia and Herzegovina must keep up with the world scientific community and contribute to human population research aimed at fighting infectious agents, in order to solve problems regarding future pandemics as efficiently as possible.

• Acknowledgment: All funding for this study was provided by ALEA Genetic Center, Sarajevo, Bosnia and Herzegovina.
• Author’s contribution: All authors were involved in preparation of this article. Final proofreading was made by the first author.
• Conflicts of interest: There are no conflicts of interest.
• Financial support and sponsorship: None

REFERENCES
Phylogenetic Changes in SARS-CoV-2 Virus in Bosnian-Herzegovinian Population Over the Period of Two Years


