sciendo
ПРИЛОЗИ. Одд. за мед. науки, XLIV 2, 2023 МАНУ
CONTRIBUTIONS. Sec. of Med. Sci., XLIV 2, 2023 MASA

10.2478/prilozi-2023-0018 ISSN 1857-9345

CIRCULATING VARIANTS OF SARS-COV-2 AMONG MACEDONIAN COVID - 19 PATIENTS IN THE FIRST YEAR OF PANDEMIC

Sanja Kiprijanovska[1], Gjorgji Bozhinovski[1], Predrag Noveski[1], Ivana Maleva Kostovska[1], Emilija Shukarova Stefanovska[1], Marija Vujovikj[1], Hristina Dichevska[1], Simona Jakovchevska[1], Natalija Jovanovska[1], Icko Gjorgoski[3], Dijana Plaseska Karanfilska[1] and Aleksandar Dimovski[1, 2]

[1] Research Center for Genetic Engineering and Biotechnology "Georgi D. Efremov", Macedonian Academy of Sciences and Arts, Skopje, RN Macedonia

[2] Center for Biomolecular Pharmaceutical Analyzes, UKIM Faculty of Pharmacy, Skopje, RN Macedonia

[3] Faculty of Natural Sciences and Mathematics, Ss Cyril and Methodius University in Skopje, Skopje, RN Macedonia

Corresponding author: Aleksandar Dimovski, PhD, Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov", Macedonian Academy of Sciences and Arts, Krste Misirkov 2, POB 428, 1000 Skopje, N Macedonia. E-mail: a.dimovski@manu.edu.mk

ABSTRACT

Genomic epidemiology has proven to be a useful tool for investigating pandemic outbreaks and tracking pathogen spread and evolution. This study describes the circulation of SARS-CoV-2 strains in N. Macedonia during a period of one year, encompassing three waves of the COVID-19 pandemic. A certain percentage (2-3%) of positive cases were continuously selected and analyzed by whole genome sequencing (WGS) technology. Using this approach, a total of 337 SARS-CoV-2 genomes were sequenced and 26 different lineages belonging to 7 clades were detected. During the first wave of the pandemic, the most dominant lineage was B.1.1, followed by B.1.1.70, which became the most dominant in the second wave. The B.1.1.7 lineage completely overpassed all other variants in the third wave. Our study strengthens the notion that the progression of COVID-19 pandemic is associated with emergence of new SARS-CoV-2 variants with increased virulence. The measure of the impact of this viral dynamic on the spread of the pandemic should be evaluated in association with other factors that might influence the transmission.

Keywords: whole-genome sequencing, SARS-CoV-2, COVID-19, Genomic epidemiology, clade, lineage

INTRODUCTION

The ongoing global pandemic of coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) started in Wuhan, China in December 2019, and since has spread around the world [1]. With more than 210 million cases and more than 4.4 million deaths (as of August 25, 2021) COVID-19 represents one of the deadliest pandemics in history (https://coronavirus.jhu.edu/). During the spread of this pandemic, SARS-CoV-2 has evolved in several major strains with increased virulence, each being associated with consecutive waves of the disease in particular geographic locations [2]. Detection of variants of the SARS-CoV-2 virus plays vital role in understanding the nature of transmission and its evolutionary mechanisms, with a primary goal to control the spread of COVID-19 and further improve the therapeutic regimens. At present (August 25, 2021), SARS-CoV-2 is classified into 8 or 19 major clades according to GISAID and NEXSTRAIN [3, 4] databases, respectively, with numerous lineages within each clade. As of May 31, 2021, based on transmissibility, immunity and infection severity, the WHO has specified

two important SARS-CoV-2 variant categories: variants of interest (VOI) and variants of concern (VOC) (https://www.who.int). Most of the data on the distribution of SARS-CoV-2 variants are from several major EU countries and the USA, whereas there is limited information on their frequency in other countries. More precise data about the global distribution of the SARS-COV-2 virus during the pandemic is crucial for investigating the outbreak and tracking the virus' spread and evolution. In this paper we describe the circulating SARS-CoV-2 variants detected by whole genome next generation sequencing of 337 isolates from N. Macedonia during a period of one year (April 2020 – May 2021), encompassing three waves of the pandemic.

MATERIALS AND METHODS

In the period of April 2020 – May 2021, 38,884 patients were referred to the Research Center for Genetic Engineering and Biotechnology "Georgi D. Efremov" (RCGEB) at the Macedonian Academy of Sciences and Arts (MASA) for SARS-CoV-2 virus testing, of which 9424 were detected as positive. A total of 337 samples (mean age 60.9±8.3 years, 139 females and 198 males) with cycle threshold (Ct) value <25 were selected for SARS-CoV-2 genome sequencing, with balanced distribution of their place of residence (Fig 1). Informed consent was obtained from all patients in accordance with the Declaration of Helsinki and the study was approved by the Ethics Committee of the Macedonian Academy of Sciences and Arts.

RNA extraction and SARS-CoV-2 detection

Nasopharyngeal swab specimens were collected from patients in sterile microbiological tubes containing specific viral transport medium (VTM) and stored/transported at 4oC until RNA extraction. RNA was isolated from 200µl swab medium using the Maccura Specimen Preparation Kit (Maccura Biotechnology Co, Chengdu, China) and eluted in final volume of 50 µl. SARS-CoV-2 RNA detection was performed with commercial kit (Maccura Biotechnology Co, Chengdu, China) and/or in-house assay using Centers for disease control and prevention (CDC) primers and procedures (CDC, https://www.cdc.gov, 2020).

Whole Genome Sequencing of SARS-CoV-2

Whole Genome Sequencing (WGS) was performed using the Respiratory Virus Oligo Panel and Illumina RNA Prep with Enrichment kits. Libraries were pooled and sequenced on the MiSeq instrument (Illumina, San Diego, CA, USA). Alignment of the sequenced reads, quality control, variant calling and generation of the consensus FASTA files was performed using Illumina's

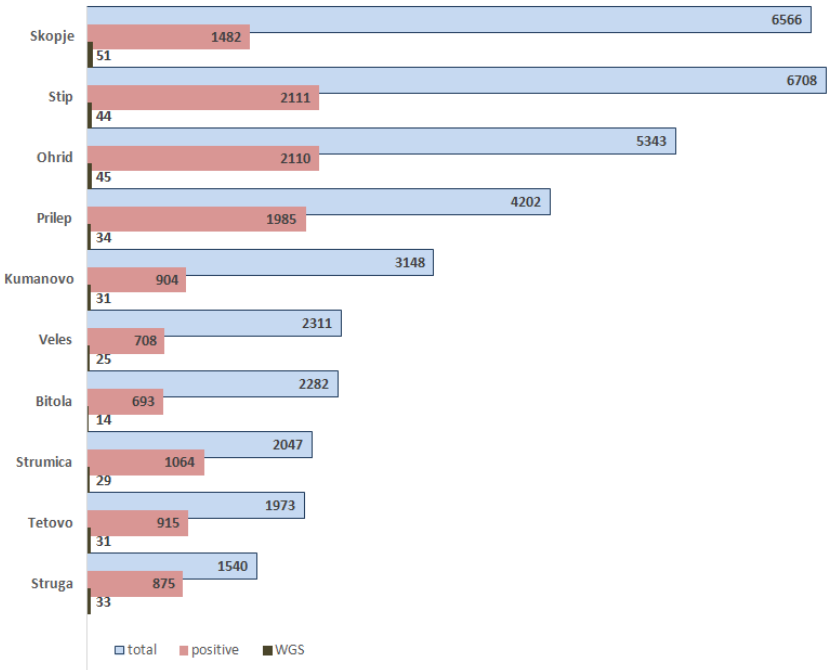


Fig 1. *The place of residence of patients included in the study. The number of patients from each city referred for diagnosis are shown in comparison with the number of SARS-CoV-2 positive patients and patients selected for NGS analysis.*

BaseSpace DRAGEN RNA Pathogen Detection v.3.5.15 App, with default settings applied. Clade assignment of the virus genomes was performed with the Nextclade tool v.0.14.2 (2021-03-30) using Nextstrain's web page (https://clades.nextstrain.org/; accessed 20 April 2021). For lineage assignment we have used web-based Pangolin COVID-19 Lineage Assigner v2.3.8 (https://pangolin.cog-uk.io/; accessed 20 April 2021). All complete virus genomes and corresponding patient metadata were submitted into GISAID database. File manipulation and graph generation was performed using R software, v.3.6.1.

RESULTS

This study includes samples referred for SARS CoV-2 analysis at the RCGEB-MASA in the period April 2020 – May 2021. The samples originated primarily from outpatient and inpatient departments from the public hospitals outside of the capitol, Skopje. In addition, samples from one COVID-19 department, a major hospice institution and several community testing points in Skopje have also been referred for diagnosis. A total of 38,884 samples were analyzed, representing 7.0% of all samples tested in N. Macedonia in the reporting period. Of these, 9424 were newly diagnosed patients with SARS-CoV-2 infection representing 9.6 % of all newly diagnosed patients in our country in the respective period. The timeline of the number of analyzed and SARS-CoV-2 positive cases shows three peaks of the pandemic: June/July 2020, November/December 2020 and March/April 2021 (Fig 2).

In order to determine the presence of different genetic variants of SARS-CoV-2, we analyzed 337 SARS-CoV-2 positive samples (3.5% of the newly diagnosed patients), using NGS. Phylogenic analysis of the NGS data classified the SARS-CoV-2 samples into seven clades (20B, 20IA/501Y.V1, 20A, 20E (EU1), 20A.EU2, 20C and 20D) and 26 lineages (Fig 3). The 20B clade was the most predominant with a frequency of 54.0%, followed by 20IA/501Y.V1 and 20A found with 26.7% and 18.9%, respectively (Fig 3). The other clades were present with a frequency of 0.3 to 3.0 %. The distribution of SARS-CoV-2 lineages showed that the most dominant lineage was B.1.1.70 with the frequency of 32.04%, followed by B.1.1.7 (26.7%), while B.1.1 and B.1.258.17 were found in 10.9% and 4.7% of sequenced genomes, respectively (Fig 3). The other 22 linages were relatively rare and were detected with a frequency of 0.3 to 3.5 %.

The temporal distribution of the most common lineages is shown in Fig 4. The data are presented in 4 periods (period 1 = April 2020 – October 2020, period 2 = November 2020 – December 2020, period 3 January 2021 – February 2021 and period 4 = March 2021 – May 2021), encompassing the three waves (periods 1, 2 and 4) and one interwave period between the second and the third wave (period 3). During period 1, the most predominant lineage was B.1.1, followed by B.1.1.70, which become the most predominant lineage in the period 2. In the period 3 a large variety of different variants were observed. The B.1.1.7 was dominant in period 4, completely suppressing all other SARS-CoV-2 variants.

The temporal distribution of the number of mutations/samples in comparison to the SARS-Co-2 referent genome (NC_045512.2) is shown in Fig 5. In general, the number of mutations/samples showed a linear increase during the pandemic from around 10 mutations at the beginning of the pandemic to >30 mutations one year later. The widest spectrum of different mutations was observed in the period 3, probably due to the greatest variety of different lineages (Fig 4). In period 4, there was a striking narrowing of the spectrum of mutations due to the predominance of only one lineage (B.1.1.7).

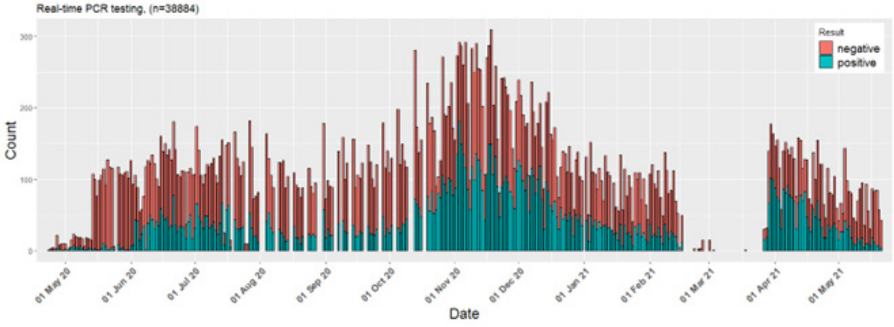


Fig 2. *Number of tested and SARS-CoV-2 virus positive cases referred to RCGEB - MASA in the period April 2020 – May 2021*

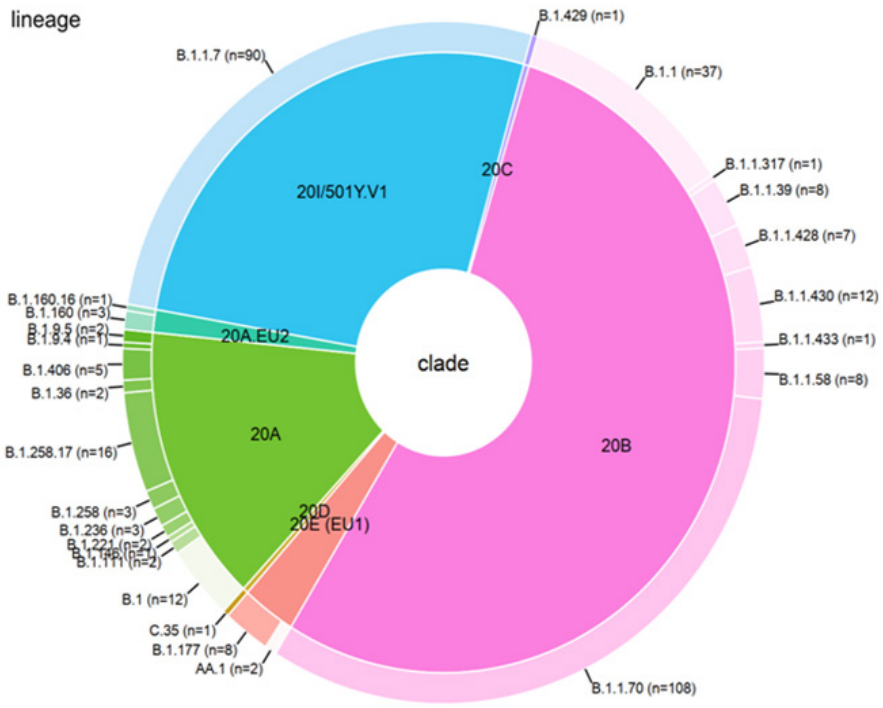


Fig 3. *Distribution of the detected SARS-CoV-2 clades/lineages*

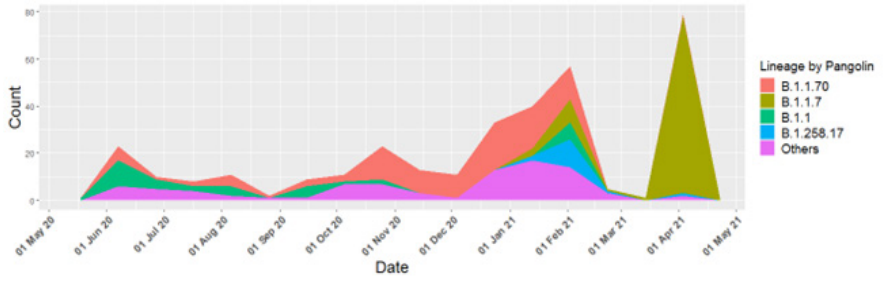


Fig 4. *Timeline distribution of SARS-CoV-2 lineages during the 4 periods of the pandemic*

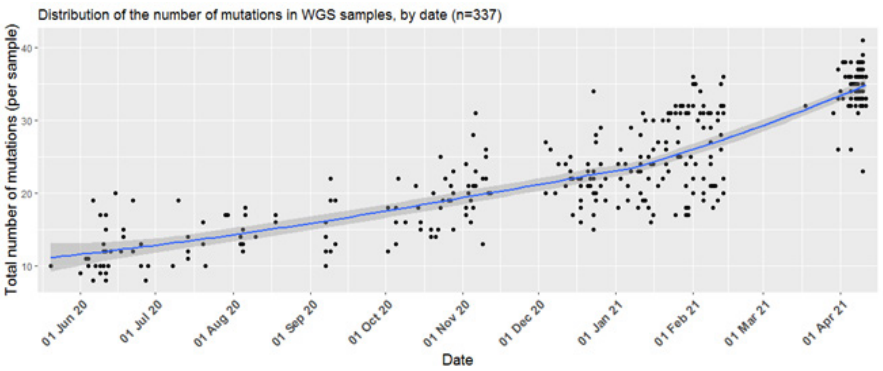


Fig 5. *Number of mutations in SARS-CoV-2 genomes detected during the course of the pandemic in N. Macedonia*

DISCUSSION

Starting from the end of February 2020, when the first case of SARS-CoV-2 infection was identified in RN Macedonia, a total of 155,304 cases and 5,423 deaths have been reported by the end of May 2021 (https://www.iph.mk). Since there were three specific rises in the number of new cases, the first year of the pandemic was categorized into three waves. This study evaluated the types and frequencies of different SARS-CoV-2 variants during the pandemic in our country. The most notable observation was that each of the waves of the epidemic was associated with an emergence of one dominant strain of the virus; B.1.1, B.1.1.70 and B.1.1.7 in the first, second and the third wave, respectively. This distribution most probably results from the combination of many factors, one of which is the increasing virulence of the new dominant strains. This assumption was most notably demonstrated with the predominance of the B.1.1.7 variant in the third wave of the epidemic, which in a relatively short period completely surpassed many other variants detected in the interwave period, including two variants that were present with a relatively high frequency (i.e., B.1.1.70 and B.1.258). The B.1.1.7 variant was first detected in southeast England in September 2020, becoming the dominant lineage in at least 114 other countries worldwide [5]. This variant has many mutations in the S-gene, particularly in the receptor binding domain, which enhances virus binding affinity to ACE2 receptor leading to higher transmissibility and virulence [6]. In line with this, it is prudent to assume that the emergence of the B.1.1.70 as the dominant strain in the second wave was due to the additional mutations gained during the epidemic. Indeed, our initial results indicate that the B.1.1.70 strain, dominant in the second wave, differs from the original strain in one mutation, Q677H which is located in the vicinity of the furin cleavage site of the S-protein [7], presumably leading to an easier cell entry for the virus (manuscript in preparation).

CONCLUSIONS

In conclusion, our study strengthened the notion that the progression of the COVID-19 pandemic is associated with the emergence of new variants with increased virulence, associated with each new wave. The measure of the impact of this viral dynamic on the spread of the pandemic should be evaluated in association with other factors that might influence the transmission, factors like public health measures, distribution of naturally acquired or vaccine induced immunity in different subpopulation groups and/or seasonality factors.

Author Contributions

Conceptualization, AD.; methodology, SK, GB, PN.; software, PN.; validation, DPK and AD.; formal analysis, SK, GB, PN; investigation, SK, GB, PN, MV, DPK and AD.; resources, AD.; data curation, PN.; writing—original draft preparation, SK, GB, PN; writing—review and editing, DPK and AD; visualization, PN.; supervision, DPK and AD

All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the funds for science of the RCGEB-MASA and donations from the local office of Hoffmann-LaRoche, ALKALOID-Skopje, EL-TA90-Skopje (local Illumina representative) and Pharmahem-Skopje.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki guidelines and the study was approved by the Ethics Committee of the Macedonian Academy of Sciences and Arts.

Informed Consent Statement

Informed consent was obtained from all patients.

Data Availability Statement

The data presented in this study are openly available in GISAID at https://www.gisaid.org/.

Acknowledgments

The authors are in debt to the patients, numerous health professionals who referred patient samples for analysis, as well as to the companies that provided financial support for this study.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

1. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical characteristics of coronavirus disease 2019 in China. New England journal of medicine. 2020;382(18):1708-20.
2. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. Frontiers in microbiology. 2020;11:1800.
3. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance. 2017;22(13):30494.
4. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-3.
5. Hodcroft E, Zuber M, Nadeau S, Crawford K, Bloom J, Veesler D, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv. Cold Spring Harbor Laboratory Press; 2020.
6. Ali F, Kasry A, Amin M. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. Medicine in drug discovery. 2021;10:100086.
7. Johnson BA, Xie X, Kalveram B, Lokugamage KG, Muruato A, Zou J, et al. Furin cleavage site is key to SARS-CoV-2 pathogenesis. BioRxiv. 2020.

Резиме

ГЕНЕТСКА ИДЕНТИФИКАЦИЈА НА ЦИРКУЛИРАЧКИТЕ SARS-COV-2 ВАРИЈАНТИ ВО ТЕКОТ НА ПРВАТА ГОДИНА ОД ПАНДЕМИЈАТА ВО РС МАКЕДОНИЈА

Сања Кипријановска[1], Ѓорѓи Божиновски[1], Предраг Новески[1], Ивана Малева Косовска[1], Емилија Шукарова Стефановска[1], Марија Вујовиќ[1], Христина Дичевска[1], Симона Јаковчевска[1], Наталија Јовановска[1], Ицко Ѓоргоски[3], Дијана Плашеска Каранфилска[1], Александар Димовски[1, 2]

[1] Истражувачки центар за генетско инженерстви и биотехнологија „Георги Д. Ефремов”, Македонска академија на науките и уметностите, Скопје, РС Македонија

[2] Центар за биомолекуларни анализи, УКИМ, Фармацевтски факултет, Скопје, РС Македонија

[3] Природно-математички факултет, УКИМ, Скопје, РС Македонија

Геномската епидемиологија претставува корисна алатка во истражувањето на пандемијата и следењето на ширењето и еволуцијата на самиот патоген. Оваа студија ја опишува циркулацијата на SARS-CoV-2 соевите во РС Македонија во период од една година, опфаќајќи три брана на пандемијата на COVID-19. Одреден процент (2–3 %) од позитивните случаи беа континуирано селектирани и анализирани со технологија за секвенционирање на целиот вирусен геном (WGS). Користејќи го овој пристап, секвенционирани беа вкупно 337 геноми на SARS-CoV-2 и беа откриени 26 различни варијанти што припаѓаат на 7 соја. За време на првиот бран на пандемијата најдоминантна варијанта беше B.1.1, по што следуваше B.1.1.70, која стана најдоминантна во вториот бран. Варијантата B.1.1.7 целосно ги потисна сите други варијанти во третиот бран. Нашата студија ја засилува идејата дека прогресијата на пандемијата на COVID-19 е поврзана со појавата на нови варијанти SARS-CoV-2 со зголемена вирулентност. Динамиката на вирусното ширење во текот на пандемијата на COVID-19 треба да се разгледува и да се асоцира и со други фактори за нејзино подетално објаснување.

Клучни зборови: целосно геномско секвенционирање, SARS-CoV-2, COVID-19, геномска епидемиологија, варијанта, соj