DOI: 10.5281/zenodo.15879875
UDC: 578.834.1:578.5(478)

GENOMIC MONITORING OF SARS-COV-2 VARIANTS IN THE REPUBLIC OF MOLDOVA

Svetlana Colac[1,2], Victor Sitnic[1,3], Olga Burduniuc[1,2]

[1] *National Agency for Public Health, Chisinau, Republic of Moldova*
[2] *"Nicolae Testemitanu" State University of Medicine and Pharmacy, Chisinau, Republic of Moldova*
[3] *Moldova State University, Chisinau, Republic of Moldova*

Summary
Objectives. The sequencing of the SARS-CoV-2 genome is essential in the process of monitoring the evolution, spread and public health impact of this virus. Since the emergence of the Omicron variant in November 2021, a number of Omicron-derived lineages (such as BA.2, BA.4/BA.5, XBB, JN.1) have continued to evolve and replace previous variants.
The aim of this paper is to analyze the clades, recombinant variants and genovariant lines of SARS-CoV-2 «Omicron» that emerged in 2024 as potential main agents of new increases in COVID-19 morbidity in the Republic of Moldova.
Methods. In the present study, Illumina technology and the MiSeq tool were used for the two-release sequencing of 71 clinical samples of SARS-CoV-2 virus, identifying the variants circulating in the Republic of Moldova between December 2023 and August 2024.
Results. The presence of 10 clades and 18 genetic lines was established. The predominant genetic lineage in the first half of 2024 was JN.1, while JN.1.11.1 and KP.2 dominated from June to August 2024. The ACE2 affinity score and nucleotide diversity of the investigated viral genomes were also determined.
Conclusions. The identified predominant variants were found to be in accordance with the data reported by the World Health Organization and research in the field.
Keywords: sequencing, SARS-CoV-2, COVID-19, Illumina, NGS, genome

Introduction

In December 2019, an outbreak of pneumonia of unknown origin was first identified in Wuhan, Hubei Province, China. Pneumonia cases were epidemiologically linked to a seafood market in Wuhan, China, caused by a novel respiratory virus [1]. This virus, now known as SARS-CoV-2, has spread rapidly worldwide, infecting more than 770 million people and causing more than 7 million deaths worldwide by October 2024 [2].

Despite WHO's announcement on May 5, 2023, that the COVID-19 pandemic had ended, the spread of the disease has not stopped and SARS-CoV-2 continues to spread through the population, creating new lineages and sublineages [3].

The new genetic variants of SARS-CoV-2 are formed mainly due to nucleotide substitutions in protein S during the natural evolution of the pathogen. These amino acid substitutions may increase the tropism of the pathogen towards the receptors of susceptible cells. Another possible mechanism for the development of new variants of SARS-CoV-2 is genetic recombination, which is possible when different variants of the pathogen circulate together in the population [4].

The potential consequences of the emergence of new genetic variants of SARS-CoV-2 are increased transmissibility of the pathogen and disease severity, decreased efficacy of diagnostic tools, prophylaxis and treatment, and the ability to cause disease in vaccinated and previously infected individuals [5].

In early 2024, the JN.1 variant (BA.2.86.1.1.1.1) replaced the XBB.1 lineage as the previously predominant variant. Subsequently, a new variant KP.2, derived from the parental variant JN.1 with the R346T, F456L mutation in the spike region, emerged. The variant has been reported in at least 28 countries, including the USA, the United Kingdom, Canada, Australia, India and Thailand. This variant is more infectious and has caused a further increase in the number of cases in many countries, including Moldova [6].

The aim of this paper is to analyze the sub-lineages of the "Omicron" genovariant of SARS-CoV-2 that emerged in 2024 as potential main agents of increased COVID-19 morbidity in the Republic of Moldova.

Materials and methods

A total of 71 samples obtained from December 2023 to August 2024 from patients hospitalized in public healthcare facilities with COVID-19 symptoms were included in the study. Samples were selected with high viral loads, specifically those with a Real-Time PCR CT value lower than 25. This cohort was chosen to study the ongoing molecular evolution of SARS-CoV-2 by whole genome sequencing. Some documented metadata, such as geographical information, vaccination status (when provided) and some hospitalization details, did not show statistical significance due to the small size of the study group.

Extraction of total RNA was performed from biological samples (nasopharyngeal and oropharyngeal exudate) stored at -70°C using Qiagen QIAamp Viral RNA extraction kits. Subsequently, genomic libraries were prepared using the commercial Illumina COVIDSeq Assay kit, which consists of indexing with barcodes, performing the PCR reaction and fragment size selection using magnetic beads. The Illumina COVIDSeq Assay kit (96 samples) includes all the reagents required for RNA to cDNA conversion, amplification, and library preparation. In addition, the modified and optimized ARTIC *v4 Primer Pools* (COVIDSeq v4 Primer Pools) were used in the library preparation process to improve viral genome coverage. The sequencing process was performed with the Illumina MiSeq Dx instrument using the MiSeq Reagent Kit v2 cartridge (300 cycles) with paired-end reads (2 × 150 nucleotides). Fluorescent images were analyzed using the Illumina base recognition platform to produce demultiplexed data in *FASTQ* format.

A total of 142 FASTQ pair-end files obtained by sequencing 71 clinical samples of SARS-CoV-2 in 2 runs were analyzed. Two samples were excluded as they did not meet the required quality criteria. Quality assessment of the raw genomic data was performed for each individual sample using fastqc [7] and aggregation of the results of all samples - with the multiqc algorithm [8]. Although, for most of the fastq files, a relatively high percentage of duplications was observed, the obtained Phred score has values above 30 and the GC percentage is around 38-40% which is specific for SARS-CoV-2 genome. The length of reads varies in the range 126-140 bp, and the number of reads per fastq file - between approx. 0.25 and 1.3 million (Figure 1).

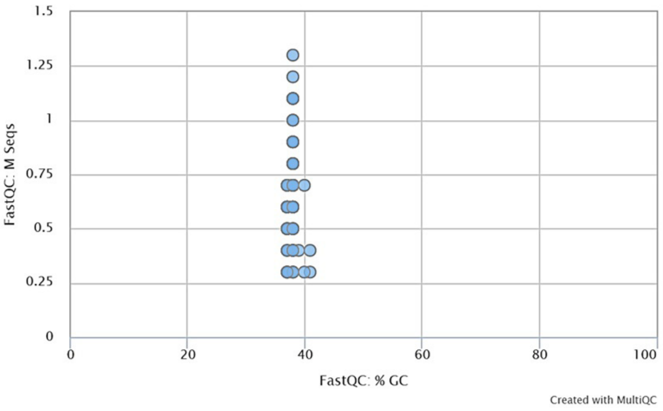

Figure 1. *Distribution of read counts and GC percentage*

Contamination of adapter sequences is low and within quality parameters. Nextclade [9] and Pangolin [10] tools were used to identify SARS-CoV-2 clades and genetic lineages in the analyzed samples. In addition to descriptive statistics (frequencies, percentages, mean Phred scores, GC content) *Shannon entropy index* was calculated to estimate nucleotide diversity across the viral genome. R language [11] and Nextstrain platform [12] were also used to create phylogenies, assess the nucleotide diversity of the 69 viral genomes included in the study and determine the affinity score for the human ACE2 receptor. The data analysis tools used in the study allowed an efficient identification of SARS-CoV-2 lineages and a robust interpretation of the viral phylogenetic patterns.

Results

Bioinformatics analysis identified 9 clades of the Omicron variant and one recombinant clade (out of 27 registered on the Nextstrain platform at the end of 2024) in the SARS-CoV-2 samples searched. Thus, the Omicron clades identified are 23A, 23B, 23D, 23H, 24A, 24B, 24C, 24E and 24G. A moderately low ACE2 affinity score was determined for clade group 23, moderate for clade 24A samples (60-70 mutations) and with a decreasing trend especially for 24C and 24E (70-80 mutations) (Figure 2).

A crucial aspect in monitoring the evolution of SARS-CoV-2 is the identification of genetic recombination events, as these can lead to the emergence of new viral strains with altered virulence and characteristics. In the analyzed samples, 3 samples with recombinant viral RNA were identified, however, no change in ACE2 affinity score was observed.

Further analysis with the Pangolin tool identified 18 genetic lines, the frequencies of which are specified in Table 1.

Table 1

Frequency of identified SARS-CoV-2 genetic lines

N/o	Identified genetic lineage	No. of samples	Frequency (%)
1	JN.1	27	39.13
2	JN.1.11.1	8	11.59
3	KP.2	7	10.14
4	BA.2.86.1	5	7.25
5	JN.1.13.1	3	4.35
6	JN.1.16	3	4.35
7	JN.1.8	2	2.9
8	KP.1.1	2	2.9
9	XBB.1.16.6	2	2.9
10	XDA	2	2.9
11	FL.15.1.1	1	1.45
12	HK.3.2	1	1.45
13	JD.1.1	1	1.45
14	JD.1.1.3	1	1.45
15	JN.1.1	1	1.45
16	JN.1.4	1	1.45
17	JN.1.5	1	1.45
18	XBB.1.16.15	1	1.45

Figure 3 shows the distribution of detected viral lineages according to the sequencing period.

The radial phylogenetic tree of all virus subvariants identified in the present study is shown in Figure 4.

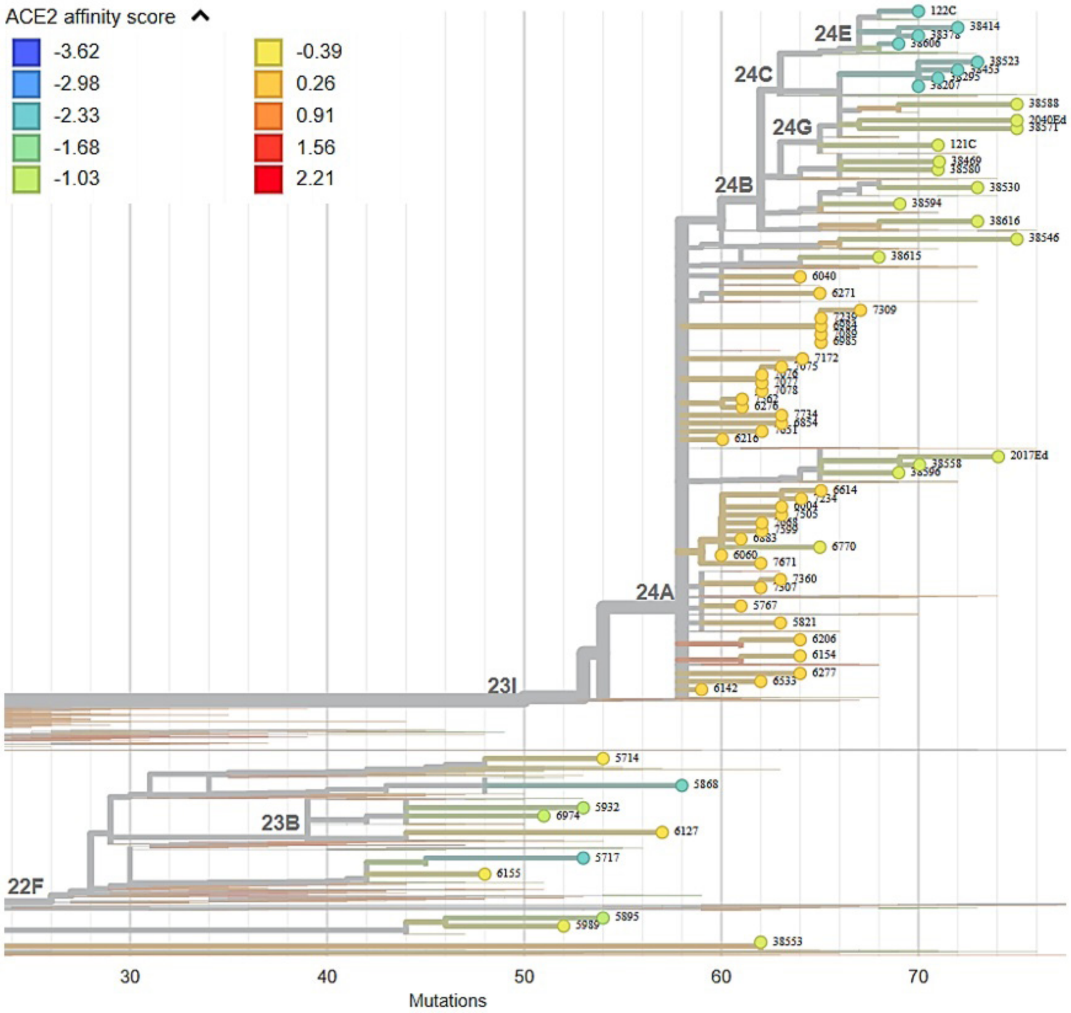

***Figure 2.** ACE2 affinity score in analyzed samples*

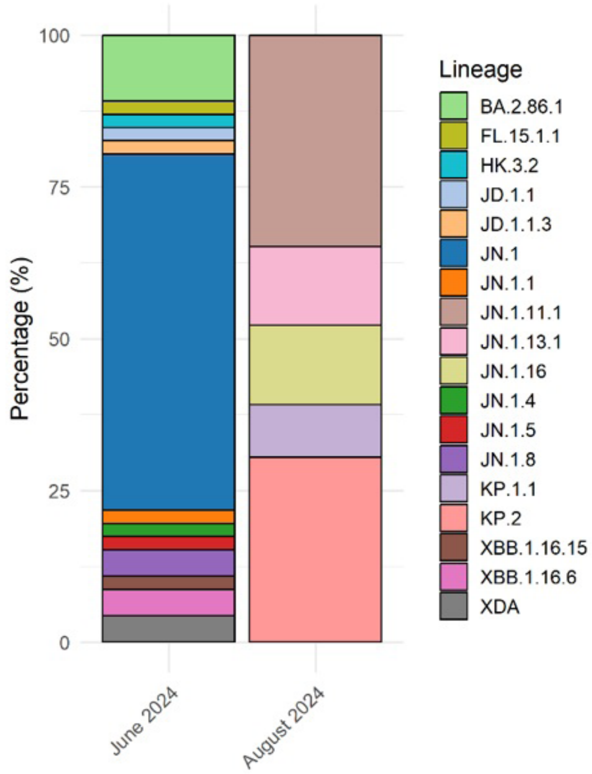

***Figure 3.** Distribution of SARS-CoV2 genetic lineages by sequencing period*

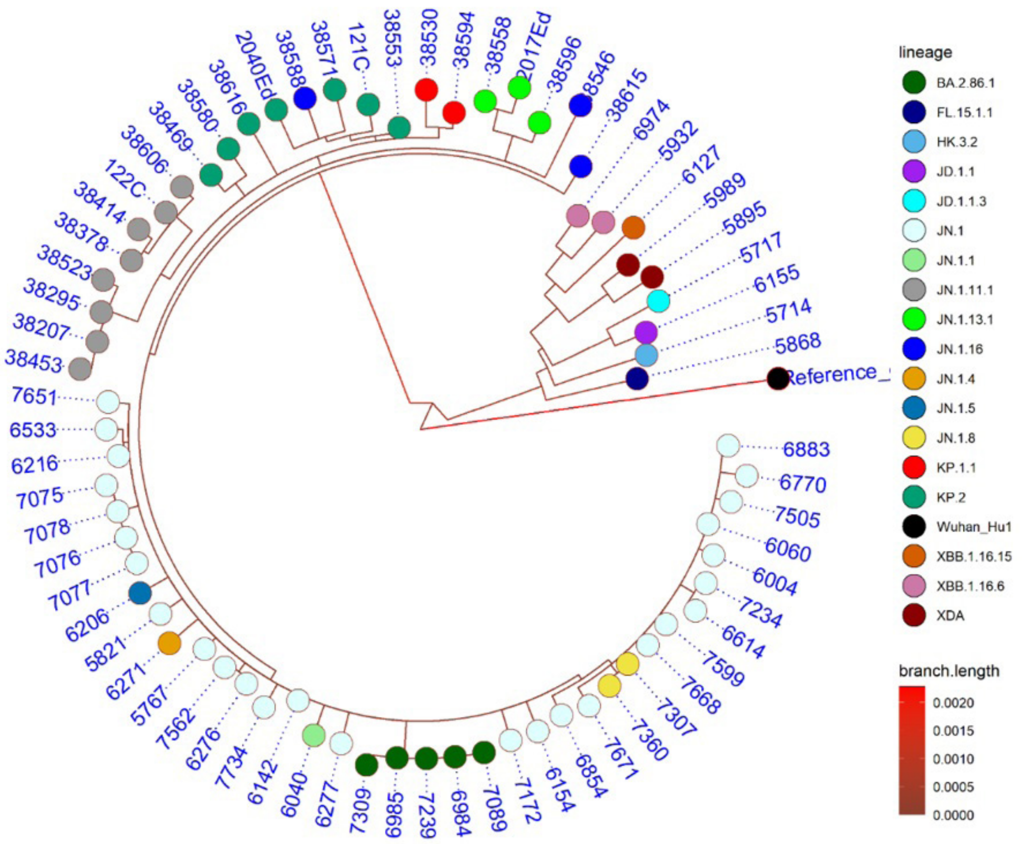

Figure 4. *Phylogenetic tree of SARS-CoV-2 lines detected in 2024*

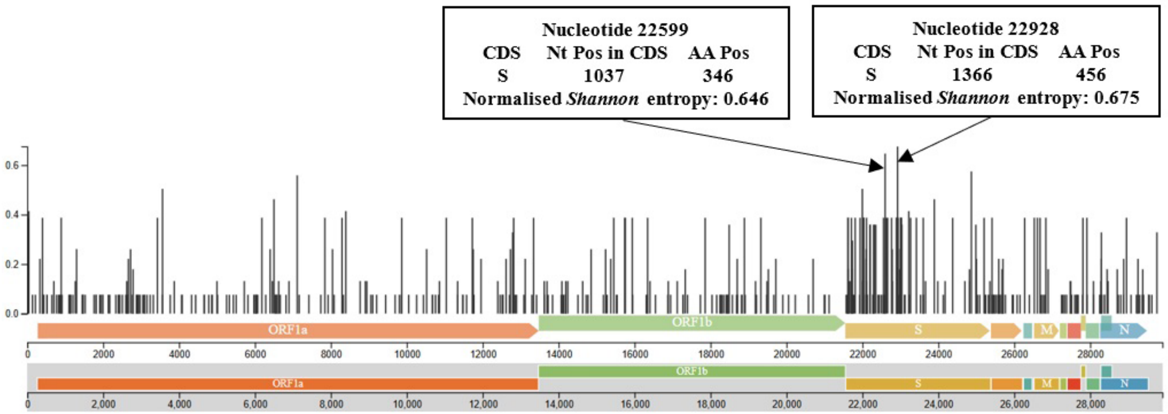

Figure 5. *Nucleotide diversity in analyzed viral genomes*

Tree branches represent evolutionary divergence, with length and color being proportional to the genetic distance between sequences. The red color of the branches implies significant genetic and evolutionary changes, while the brown color suggests greater similarity between sequences.

The *Shannon index* of nucleotide diversity in the 69 sequenced viral genomes is shown in Figure 5. The X-axis shows the genomic position and the Y-axis the nucleotide diversity.

It is noticed that the first two positions with the highest nucleotide diversity are in the coding region (CDS) of the Spike protein. The substitution positions are in genomic regions 22928 and 22599, and at the amino acid level: 456 and 346, respectively. High diversity regions represent mutation-prone hotspots in the genome and are driven by evolutionary pressures such as immune evasion, whereas low diversity regions are more highly conserved and are usually involved in virus replication and reproduction.

Discussions

SARS-CoV-2 genetic variations are grouped into clades, also known as subtypes, genotypes or phylogenetic groups. The clades reflect the genetic diversity of the virus and allow the virus to be tracked as it adapts to its host. Nextstrain uses a systematic and standardized approach to clade naming, which facilitates the comparison and understanding of viral evolution on a global scale and provides insight into the molecular dynamics of the virus. The naming strategy has been continuously optimized to keep pace with the changing dynamics of SARS-CoV-2 and the rapid accumulation of mutations. The clades obtained in the present study show a maximum of 70-80 cumulative mutations, including deletions, insertions or nucleotide substitutions. These genetic changes are essential for virus adaptation to novel conditions and can influence important virus traits. For example, mutations in functional regions can alter the affinity score for the human ACE2 receptor thereby influencing transmissibility and disease severity.

Another important aspect of the molecular dynamics of SARS-CoV-2 is genetic recombination, which can occur when a cell is simultaneously infected by two or more viral strains. Under these conditions, the exchange of genetic material between strains can lead to the formation of chimeric viruses, which contain RNA fragments from several viral subtypes. These may have altered functional characteristics related to transmissibility, pathogenicity or response to vaccines and antiviral treatments [13]. The investigations identified 3 recombinant samples that did not show changes in ACE2 affinity score. This suggests that, although genetic recombination may contribute to the genomic diversity of the virus, not all recombinants result in major functional changes. However, continued monitoring of these phenomena is essential for understanding the evolution of the virus and identification of potential risks associated with the emergence of novel strains.

Figure 3 shows that between December 2023 and June 2024, the predominant genetic lineage was JN.1, which is consistent with data published by the World Health Organization (WHO) for the period [14]. The JN.1 lineage dominated in prevalence and geographical distribution, being identified in the majority of samples analyzed globally. This reflects its increased adaptability and evolutionary advantage in the context of viral circulation at that time. Subsequently, in the summer of 2024, the highest prevalence was recorded for the genetic lines JN.1.11.1 and KP.2. Importantly, KP.2 is a subvariant of the FLiRT genetic group and is derived from the JN.1.11.1 line, which in turn is descended from the JN.1 line. This is an example of the genetic diversification of SARS-CoV-2, with the emergence of subvariants that have been able to better adapt to selective environmental pressures, including those imposed by population immunity and public health interventions. Phylogenetic analysis shows that FLiRT group subvariants, such as KP.1.1 and KP.2, are localized in the same cluster, which confirms the genetic proximity between them and the predominant line in the first part of 2024, JN.1, forms a distinct cluster that includes almost half of the samples investigated in the present study. This distribution highlights the initial dominance of the JN.1 lineage and its succession over time by the more specialized subvariants in the FLiRT cluster. FLiRT is a term used to describe a family of SARS-CoV-2 genetic lines that exhibit 2 mutations in the Spike protein that were not observed in JN.1. Thus, at the protein level, phenylalanine (F) at position 456 is replaced by leucine (L) and arginine (R) at position 346 - by threonine (T). These changes have significant effects on the properties of the virus, in particular its ability to avoid neutralization by antibodies. Thus, mutations associated with the FLiRT group contribute to increased resistance to the host immune response, which may translate into increased transmissibility and reduced efficacy of existing vaccines. The FLiRT subvariants demonstrate how SARS-CoV-2 continues to evolve through selective pressure, favoring genetic lineages that are able to evade the immune response and remain evolutionarily competitive. These molecular dynamics confirm the need for constant monitoring of new sub-variants to anticipate possible changes in transmissibility, pathogenicity or vaccine susceptibility. The continued adaptability of the virus underlines the importance of genomic monitoring of the virus and the integration of the resulting data into overall public health strategies.

Conclusions

The sequencing of 71 clinical samples of SARS-CoV-2 identified 10 clades and 18 genetic lineages, with the predominant variants in line with World Health Organization data and trends observed globally. The predominant genetic lineage in the first half of 2024 was JN.1, while JN.1.11.1 and KP.2 prevailed from June to August 2024. Three recombinant viral RNA samples were identified in which no changes in ACE2 affinity score were observed. The study also highlights high nucleotide diversity in the Spike protein coding region, particularly at genomic positions 22928 and 22599, confirming the intense evolutionary pressures on this region.

Bibliography

1. Ciotti M, Ciccozzi M, Terrinoni A, Jiang WC, Wang CB, Bernardini S. The COVID-19 pandemic. Crit Rev Clin Lab Sci. 2020;57(6):365-388. doi:10.1080/10408363.2020.1783198
2. Souza UJB, Spilki FR, Tanuri A, Roehe PM, Campos FS. Two Years of SARS-CoV-2 Omicron Genomic Evolution in Brazil (2022-2024): Subvariant Tracking and Assessment of Regional Sequencing Efforts. Viruses. 2025;17(1):64. doi:10.3390/v17010064
3. Burki T. WHO ends the COVID-19 public health emergency. Lancet Respir Med. 2023;11(7):588. doi:10.1016/S2213-2600(23)00217-5
4. Centers for Disease Control and Prevention. CDC COVID-19 Science Briefs. Atlanta, GA: US Department of Health and Human Services; 2020–2023. https://www.ncbi.nlm.nih.gov/books/NBK570439/
5. McKie R. Omicron: What do we know about the new Covid variant? The Guardian. December 5, 2021. https://www.theguardian.com/world/2021/dec/05/omicron-what-do-we-know-about-the-new-covid-variant (archived version retrieved September 4, 2024)
6. Shanmugaraj B. Ever-evolving SARS-CoV-2: Latest variant KP.2 is on the rise. Asian Pac J Trop Med. 2024;17(6):241-242. doi:10.4103/apjtm.apjtm_341_24

7. Andrews S. FastQC: A quality control tool for high throughput sequence data. Version 0.12.1. Babraham Bioinformatics; 2010. [Software]. Installed on Ubuntu 24.04.
8. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354 (Version 1.0.dev0, installed on Ubuntu 24.04)
9. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling, and quality control for viral genomes. J Open Source Softw. 2021;6(67):3773. doi:10.21105/joss.03773
10. O'Toole Á, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7(2):veab064. doi:10.1093/ve/veab064
11. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. https://www.R-project.org
12. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-4123. doi:10.1093/bioinformatics/bty407
13. Garcia I, Bråte J, Fossum E, et al. Recombinant SARS-CoV-2 Delta/Omicron BA.5 emerging in an immunocompromised long-term infected COVID-19 patient. Sci Rep. 2024;14(1):25790. doi:10.1038/s41598-024-75241-3
14. World Health Organization. Updated risk evaluation of JN.1. April 15, 2024. https://www.who.int/docs/default-source/coronaviruse/15042024_jn1_ure.pdf

Received – 29.01.2025, accepted for publication – 03.03.2025

Corresponding author: Svetlana Colac, e-mail: svetlana.colac@ansp.gov.md

Conflict of interest Statement: The authors report no conflicts of interest in this work.
Funding Statement: The authors report no financial support.
Citation: Colac S, Sitnic V, Burduniuc O. Genomic monitoring of SARS-CoV-2 variants in the Republic of Moldova. Arta Medica. 2025;94(1):11-16.