



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Isolation and phylogenetic analysis of SARS-CoV-2 variants collected in Russia during the COVID-19 outbreak



Liubov Kozlovskaya^{a,b,*}, Anastasia Piniava^a, Georgy Ignatyev^a, Alexey Selivanov^a, Anna Shishova^{a,b}, Anastasia Kovpak^a, Ilya Gordeychuk^{a,b}, Yury Ivin^a, Anastasia Berestovskaya^c, Egor Prokhortchouk^d, Denis Protsenko^c, Mikhail Rychev^d, Aydar Ishmukhametov^{a,b}

^a FSBSI "Chumakov Federal Scientific Center for Research and Development of Immune and Biological Products of the Russian Academy of Sciences" (FSBSI "Chumakov FSC R&D IBP RAS"), Moscow 108819, Russia

^b Sechenov Moscow State Medical University, Moscow 119991, Russia

^c GKB No. 40 DZM, Moscow 129301, Russia

^d NRC "Kurchatov Institute", Moscow 123182, Russia

ARTICLE INFO

Article history:

Received 8 May 2020

Received in revised form 14 July 2020

Accepted 19 July 2020

Keywords:

SARS-CoV-2

COVID-19

Epidemiology

Phylogenetic analysis

Mutation

Russia

ABSTRACT

Objectives: The outbreak of coronavirus disease 2019 (COVID-19) started in December 2019 in China and then spread worldwide over the following months, involving 188 countries. The objective of this study was to determine the molecular epidemiology of the COVID-19 outbreak in Russia.

Methods: In this study, two severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains were isolated and genetically characterized. A phylogenetic analysis of all available Russian sequences was then performed and these were compared to the epidemiological data on COVID-19 incidence to evaluate the molecular epidemiology and pattern of virus spread in the territory of Russia.

Results and conclusions: Whole genome analysis of the isolates obtained in this study and 216 others isolated in Russia revealed a set of seven common mutations when compared to the original Wuhan virus, including amino acid substitutions in spike protein S and nucleoprotein N, possibly affecting their properties. Phylogenetic analysis of all Russian sequences and 8717 sequences from other countries showed multiple importations of the virus into Russia, local circulation, and several patterns of virus spread.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

On March 11, 2020 the World Health Organization (WHO) characterized coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as a pandemic (WHO, 2020a). The outbreak started as a cluster of pneumonia cases in late December 2019 in Wuhan, Hubei Province, China (WHO, 2020b); it has now reached the magnitude of over 12 million confirmed cases in 188 countries worldwide (Dong et al., 2020; (COVID-19 Dashboard, 2020).

Coronaviruses belong to the *Orthocoronavirinae* subfamily of the *Coronaviridae* family in the order *Nidovirales* (ICTV Taxonomy, 2020). The subfamily includes four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. The two epidemiologically important previously known human pathogens severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are betacoronaviruses (Park, 2020). Genome sequencing has revealed SARS-CoV-2 to be a new betacoronavirus closely related to SARS-CoV, both belonging to the *Sarbecovirus* subgenus (Zhu et al., 2020).

The symptoms of COVID-19 are rather similar to those of diseases caused by other coronaviruses. The most common features are fever, cough, and fatigue. In the most severe cases, the respiratory infection progresses to pneumonia complicated by acute respiratory distress syndrome, resulting in death (Rabaan et al., 2020).

* Corresponding author at: Laboratory of Poliomyelitis and Other Enterovirus Infections, WHO RRL Polio, FSBSI "Chumakov FSC R&D IBP RAS", Moscow 108819, Russia.

E-mail address: lubov_i_k@mail.ru (L. Kozlovskaya).

To gain further understanding of the molecular epidemiology of the COVID-19 outbreak in Russia, this study characterized the full genome sequences of two SARS-CoV-2 strains isolated from patients and mapped them to the epidemiological timeline of the outbreak in Russia, with a focus on Moscow; a phylogenetic analysis of 217 SARS-CoV-2 genome sequences obtained during the course of the epidemic in Russia was then performed.

Materials and methods

Epidemiological data collection and analysis

Epidemiological data on COVID-19 cases registered in Russia were retrieved from daily reports of the Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor, <https://www.rospotrebnadzor.ru/>). The statistical analysis was performed using Microsoft Excel 2013.

Patients, sample collection, and virus isolation

The study sample comprised patients admitted to the GKB No. 40 DZM Hospital, Moscow, Russia with a fever (38–39 °C) and catarrhal symptoms. On the date of admission, nasopharyngeal swab samples were collected and tested for SARS-CoV-2 RNA by RT-PCR (Vector, Russia).

PCR-positive nasopharyngeal swab samples were used for virus isolation. The original specimens were diluted 1:1 with Eagle minimum essential medium (EMEM) with doubled amino acids and vitamins supplemented with 32 µg/mL trypsin (FSBSI “Chumakov FSC R&D IBP RAS”, Russia); the dilutions were added to 24-h Vero cell confluent monolayers and incubated for 1 h at room temperature. Culture medium containing antibiotics and 16 µg/mL trypsin was then added. The cells were incubated at 37 °C and observed daily for signs of a cytopathic effect (CPE). CPE signs were observed from day 4 post infection. After 6 days, cell culture suspensions were collected and examined for the presence of viral RNA by POLYVIR SARS-CoV-2 (Lytech, Russia).

Sequencing

A total of 100 µL of infected cell culture suspension was used for RNA extraction using the RNeasy Mini Kit (Qiagen). RNA was then fragmented using the RNA fragmentation module (NEB) and used for library preparation with NEBNext UltraII RNA Library Prep Kit for Illumina (NEB). Sequencing was performed on a HiSeq1500 (Illumina) generating over 10 million of 250 nucleotide reads. Illumina reads were trimmed and mapped on the reference sequence (GenBank ID [NC_045512.2](#)) with Bowtie2 (Langmead and Salzberg, 2012); single nucleotide variations (SNVs) were called using SAMTools (Li et al., 2009). Additionally, mapped reads were assembled with Abyss 2.2.4. Illumina sequencing and the data analysis were performed by Genoanalytica LLC (<https://genoanalytica.ru/>).

Mutations and phylogenetic analysis

Complete genome (>29 000 nt) high coverage SARS-CoV-2 sequences were retrieved from the GISAID database (accessed June 7, 2020, about 30 000 sequences). Sequences were sorted by collection date; every third sequence and all 217 Russian sequences were collected for analysis. A total of 8934 sequences were used for further analysis with the Nextstrain framework (Hadfield et al., 2018; Sagulenko et al., 2018) to estimate the phylogeny and molecular epidemiology of the virus variants in Russia. The results can be run from <https://github.com/AlexSelivanoff/sars-cov-2-russia/tree/master/>.

For the mutations analysis, 218 Russian sequences and the reference sequence (GenBank ID [NC_045512.2](#)) were aligned with Clustal Omega (Sievers et al., 2011; Sievers and Higgins, 2018) and analyzed with MEGA X (Kumar et al., 2018). Amino acid substitutions were analyzed to investigate their impact on protein structure and function (Pommié et al., 2004).

Ethics statement

Written informed consent was obtained from all subjects or their legal representatives by the primary clinical site.

Results

Epidemiological data analysis

The first confirmed COVID-19 cases were registered in Russia in late January 2020, but a steady daily increase in the number of newly registered cases started in mid-March. About a half of all cases were registered in Moscow (Figure 1A). Moscow was the first region to implement preventive measures, starting with 14 days of home isolation for citizens arriving from affected territories (March 5, 2020). The following were then gradually implemented: (1) a ban on public events with >5000 participants (March 10, 2020) and then >50 participants (March 16, 2020); (2) transfer of all students to remote education (March 16–21, 2020); (3) home isolation for the elderly (65+ years) and those in risk groups with chronic conditions (March 26, 2020); (4) shutdown of most public places, including parks (March 28, 2020); (5) social distancing and overall isolation at home (March 30, 2020). The preventive measures were prolonged several times (until May 1, 2020 on April 2, 2020; until May 11, 2020 on April 28, 2020; until May 30, 2020 on May 7, 2020; until June 14, 2020 on May 27, 2020) (Mayor of Moscow website, 2020).

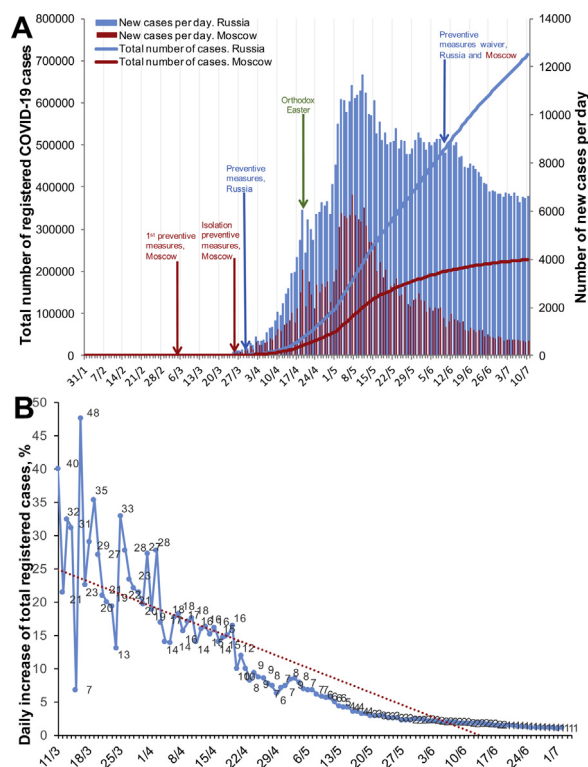


Figure 1. Dynamics of COVID-19 case registrations in Russia. (A) Total number and daily increase in cases in Moscow and Russia overall. (B) Daily increase in the total registered cases with the linear regression trend line.

Countrywide measures aimed at preventing the spread of SARS-CoV-2 infection were implemented on March 30, 2020, as the total number of reported confirmed cases was 1840. These measures included the official declaration of non-working days and social distancing ([Decree of the President of the Russian Federation, 2020](#)).

By April 6, 2020, the average daily rate of decrease in the proportion of reported confirmed cases had stabilized at 0.43% ($R^2 = 0.85$) ([Figure 1B](#)). With the current dynamics of new case registrations, we expected a decline in infection incidence to less than 1% daily by mid-May. However, an increase in overall daily registered cases was observed at the beginning of May, which could be attributed to violations of the isolation measures during Orthodox Easter (April 19, 2020). Eventually, the number of daily registered cases stabilized at 8000–9000 per day in mid-May and decreased slowly. In Moscow, the infection rate decreased faster, reaching 2000 cases per day by the end of May and 1000 cases per day by mid-June.

As the daily number of registered cases continued to decrease, the preventive measures were gradually waived: isolation measures were abolished for all citizens on June 9, 2020; cafes, restaurants, museums, etc. were reopened on June 16, 2020; sports centers, etc. were reopened on June 23, 2020 ([Mayor of Moscow Website, 2020](#)). However, the functioning of these facilities has been restricted substantially in accordance with the recommendations of the Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor, https://www.rospotrebnadzor.ru/region/korono_virus/punkt.php), including the restriction of the number of people gathering, use of personal protective equipment (PPE; masks and gloves) in public places, etc. The daily number of cases continues to decrease overall in Russia, with the tendency for a faster decrease in Moscow, being at around 600 cases per day in July.

Virus culture and whole genome sequencing

PCR-positive nasopharyngeal swab samples from patients admitted to the hospital in Moscow with fever (38–39 °C) and a cough were seeded on Vero cells. CPE signs were observed from day 4 post infection in some cultures (11/39). The culture suspensions were collected on day 6 post infection; some were discarded due to bacterial or fungal contamination (12/39). CPE-negative cultures were used for blind passage and all of them showed signs of CPE in the second passage (16/16). Cell culture suspensions were examined for the presence of viral RNA by RT-PCR.

Two positive culture suspensions with the highest cycle threshold (Ct) values were used for high-throughput sequencing (HTS). Sequences were submitted to the GISAID database as EPI_ISL 428851–2. Case descriptions are presented in [Table 1](#). Both

patients were admitted on April, 1 2020. In both cases, computed tomography (CT) showed polysegmental pneumonia. The first patient with mild symptoms was released for home treatment; the second patient was admitted, transferred to the intensive care unit on days 6–7, and discharged 15 days later with negative nasopharyngeal swabs and in a recovering condition.

The isolated SARS-CoV-2 variants (25 and 35) contained eight nucleotide substitutions each: seven common and one unique each, without any additional heterogenic positions or deletions/insertions. Five of the seven shared nucleotide mutations in the protein-coding region led to four amino acid substitutions ([Table 2](#)).

Mutations and phylogenetic analysis of SARS-CoV-2 variants isolated in Russia

Russian SARS-CoV-2 sequences were dated from March 2 to April 23, 2020, although most of the sequences represented samples isolated in April (131/218) and in Saint-Petersburg (135/218).

The overall viruses collected in Russia contained 178 nucleotide mutations, most of which were non-synonymous (108 out of 172 in the coding region). Comparison of the mutated amino acid properties (hydropathy, volume, chemical, physico-chemical properties, charge, polarity, and H-bond donor/acceptor) showed that at least 58 of the 108 amino acid substitutions were non-conservative and were most likely to affect protein structure ([Supplementary Material Table S1](#)). Most non-synonymous substitutions were found in the sequences of nucleoprotein N.

Overall, 55 mutations were found in more than one sequence and in different combinations ([Figure 2](#)).

Most Russian sequences contained the combination of four mutations $C_{241} \rightarrow T$, $C_{3037} \rightarrow T$, $C_{14408} \rightarrow T$, and $A_{23403} \rightarrow G$ (213/218) and the seven mutations identified in the cases observed in our study, i.e. the four mentioned above accompanied by $G_{28881} \rightarrow A$, $G_{28882} \rightarrow A$, and $G_{28883} \rightarrow C$ (144/218) ([Figure 2](#) and [Supplementary Material Table S2](#)). Four of these nucleotide substitutions resulted in amino acid mutations with drastically changed properties: Pro4715Leu in RNA-dependent RNA polymerase, Asp614Gly in spike protein S, and Arg203Lys and Gly204Arg in nucleoprotein N.

Only four sequences did not contain common mutations: a sample collected on March 20 in Saint-Petersburg (EPI_ISL 428,883), two samples collected on March 23 in Kabardino-Balkaria (two identical sequences, EPI_ISL 428,893 and 428,898), and one sample collected on March 24 in Saint-Petersburg (EPI_ISL 428,897). Sequences from Kabardino-Balkaria clustered with sequences collected mostly in India, Australia, and Saudi Arabia in March, but originating from Shanghai viruses collected in late January ([Figure 3A](#)).

Table 1
Clinical features of the studied cases.

| Case | Virus isolate | Sex | Age, years | Contact(s) with infected | Disease onset | First symptoms | Hospital admission | Symptoms at admission | CT | Treatment |
|------|---------------|-----|------------|--------------------------|----------------|---|--------------------|---|---|--|
| 1 | 25 | F | 42 | March 23, 2020 | 30 March, 2020 | Weakness, cough with difficult expectoration, T up to 38 °C | April 1, 2020 | Weakness, cough with difficult expectoration, T 36.2 °C | Bilateral polysegmental pneumonia | Discharged Hydroxychloroquine, azithromycin |
| 2 | 35 | M | 44 | March 11, 2020 | March 19, 2020 | Sore throat, cough with moderate expectoration, T up to 39 °C | April 1, 2020 | T 38.5 °C, cough | Polysegmental pneumonia; ground glass lesions; multiple chaotic foci; infiltrations | Hospitalized |

F, female; M, male; CT, computed tomography; T, temperature.

Table 2
Nucleotide and amino acid mutations in the genomes of SARS-CoV-2 variants described in this study as compared to the Wuhan **NC_045512.2** variant.

| Genome region | Encoded protein | Nucleotide position | Wuhan, China | Moscow, Russia | | Moscow, Russia | |
|---------------|----------------------------|---------------------|--------------|----------------|--------------------------|----------------|--------------------------|
| | | | NC_045512.2 | EPI_ISL_428851 | EPI_ISL_428852 | NT | AA |
| 5'UTR | – | 241 | C | T | N/A | T | N/A |
| ORF1ab | nsp3 (PL ^{Pro}) | 3037 | C | T | Phe ₉₂₄ | T | Phe ₉₂₄ |
| | | 3373 | C | A | Asp ₁₀₃₆ →Glu | C | – |
| | nsp5 (3CL ^{Pro}) | 10,969 | C | C | – | T | Phe ₃₅₆₈ |
| | nsp12 (RdRP) | 14,408 | C | T | Pro ₄₇₁₅ →Leu | T | Pro ₄₇₁₅ →Leu |
| ORF2 | S | 23,403 | A | G | Asp ₆₁₄ →Gly | G | Asp ₆₁₄ →Gly |
| ORF9a | N | 28,881 | G | A | Arg ₂₀₃ →Lys | A | Arg ₂₀₃ →Lys |
| | | 28,882 | G | A | – | A | – |
| | | 28,883 | G | C | Gly ₂₀₄ →Arg | C | Gly ₂₀₄ →Arg |
| | | – | G | – | – | – | – |

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; N/A - not applicable.

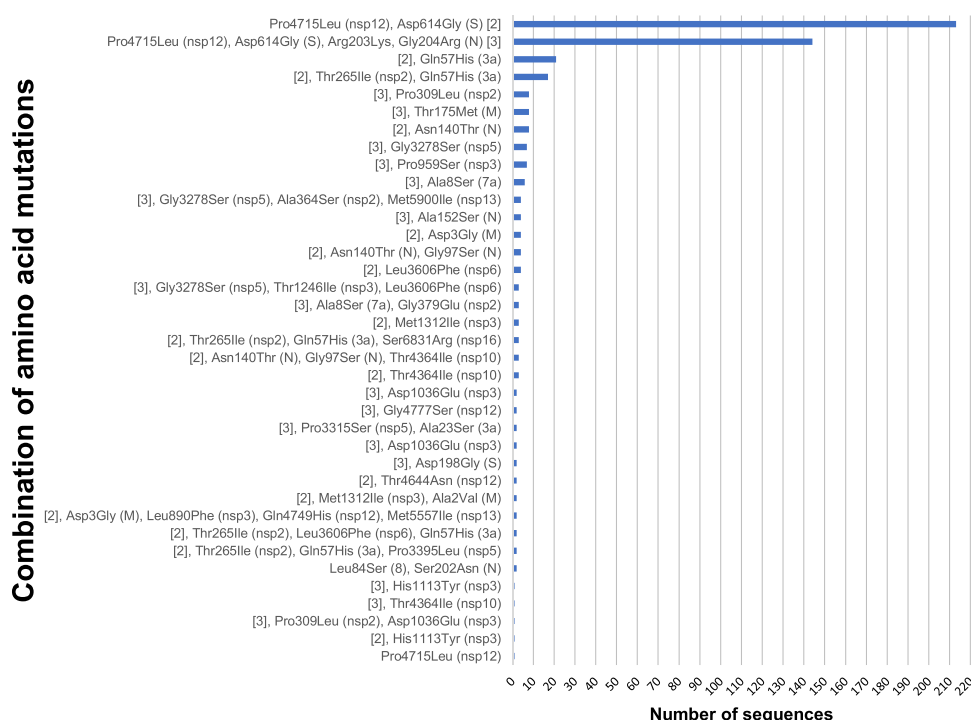


Figure 2. Combinations of amino acid mutations acquired by SARS-CoV-2 variants isolated in Russia ($N = 218$). The mutations analyzed could be found in more than one sequence. The mutated protein is shown in parentheses. The most common combinations, Pro4715Leu (nsp12), Asp614Gly (S) and Pro4715Leu (nsp12), Asp614Gly (S), Arg203Lys, Gly204Arg (N) are shortened as [2] and [3], respectively.

Among the other Russian sequences, two different groups with two common amino acid mutations Pro4715Leu (nsp12) and Asp614Gly (S), with or without N protein mutations (Arg203Lys and Gly204Arg) were observed (Figure 2). Isolates from both groups could be found throughout the isolation period (March–April) in different parts of the country. Most interestingly, some sequences from these two groups shared other common mutations: for example, sequences EPI ISL 427,321 and 427,307 contain mutations Pro4715Leu (nsp12) and Asp614Gly (S) accompanied by Asn140Thr (N), and sequences EPI ISL 450,258 and 450,260 contain the same three but with N protein mutations as well (Arg203Lys and Gly204Arg).

Another interesting set of mutations accompanying Pro4715Leu (nsp12) and Asp614Gly (S) is Thr265Ile (nsp2) and Gln57His (3a). Overall, 17 isolates containing these mutations were isolated in different parts of the country (from Moscow to Sakhalin) from March 13, to April 10, 2020.

Phylogenetic analysis showed that most of the Russian sequences clustered with the European ones, with signs of multiple introductions of the virus at different time-points from multiple locations. However, we did not find signs of direct importations from China (Figure 3B). We can see the signs of internal circulation of imported viruses: China→UK→USA→Krasnodar region→Orenburg region, as well as signs of SARS-CoV-2 circulation and mutations acquisition in Saint-Petersburg (Figure 3C). Moreover, based on the sequences available, the circulation start date for Saint-Petersburg can be estimated as the first days of March (Figure 3C).

Discussion

Epidemiological data

The COVID-19 epidemic started over 5 months ago and still poses a huge threat to public health all over the world. The first

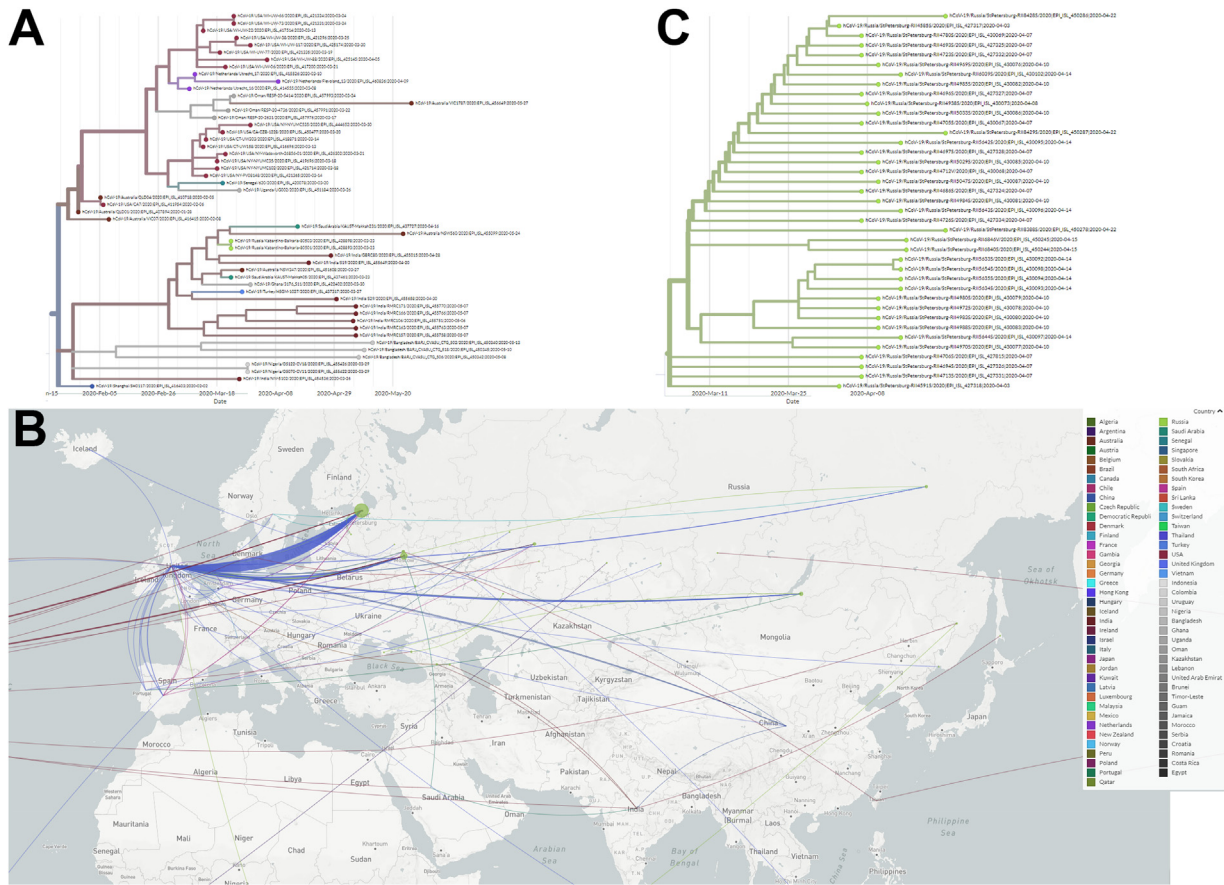


Figure 3. (A) Amplified part of the phylogenetic analysis of SARS-CoV-2 complete genome sequences, isolated in Russia and other countries (N = 8934 genome sequences, retrieved from GISAID) with sequences from Kabardino-Balkaria (shown in grass green). (B) Geographical spread of the virus in the territory of Russia. Sequences attributed to the regions were mapped to the region capital. Moscow and Moscow region are treated as different regions. (C) Amplified part of the tree with a cluster of sequences from Saint-Petersburg (shown in grass green). The tree and map were built using the Nextstrain framework.

cases were registered in Russia in the last days of January 2020, but it took another month for SARS-CoV-2 to cause a substantial number of cases and spread all over the country. When the WHO declared the pandemic of COVID-19, Moscow as the most affected city started the implementation of preventive measures, which were then applied on a countrywide scale on March 30, 2020. By April 6, 2020, the average daily rate of decrease in the proportion of reported confirmed cases had stabilized at 0.43% ($R^2 = 0.85$), which can be accepted as the result of the implementation of the preventive measures. With the observed dynamics of new case registrations, we expected a decline in infection incidence to less than 1% daily cases by mid-May. This decline was significantly slowed down following mass gatherings of people on national holidays (first week of May), despite the social distancing policies. As an example, the May peak of infection incidence (May 1–3, 2020) started 11 days after Orthodox Easter (April 19, 2020), associated with church visits, etc., despite the ‘stay home’ recommendations. The 11–14-day period corresponds to the estimated time from infection to diagnosis.

At the beginning of June, the number of new daily cases showed a steady decrease, which led to the gradual ending of the preventive measures. The registered number of cases in Russia continued to decline for the next month and stabilized at 6000–7000 cases daily. In Moscow, the decline was more rapid and the daily case number reached 600 in early July.

Analysis of the mutations identified in the Russian sequences

Mutations identified in SARS-CoV-2 sequences originating from Russia showed a low mutation rate in comparison with the reference sequence, which is not surprising for coronaviruses. However, the analysis revealed a high rate of non-synonymous mutations resulting in non-conservative substitutions. This observation may serve as indirect evidence of the intensive circulation of the virus in the human population and its adaptation to new host.

A set of seven nucleotide mutations was identified, which was presented in most Russian sequences. Four of these nucleotide substitutions resulted in amino acid mutations with drastically changed biochemical properties: Pro4715Leu in RNA-dependent RNA polymerase, Asp614Gly in spike protein S, and Arg203Lys and Gly204Arg in nucleoprotein N. These mutations have been reported for European sequences previously elsewhere (Stefanelli et al., 2020).

Mutation C₂₄₁→T was shown to co-evolve with C₃₀₃₇→T, C₁₄₄₀₈→T, and A₂₃₄₀₃→G mutations (Yin, 2020). These four mutations were abundant among European virus isolates, potentially causing a more severe course of COVID-19. It has been hypothesized that these mutations might increase the transmissibility of the virus.

Common nucleotide mutation C₁₄₄₀₈→T leading to amino acid substitution Pro4715Leu in RNA-dependent RNA polymerase was

highly presented in the Russian sequences, although the significance of this particular mutation is not well understood. It is supposedly situated at an interface domain, a surface structure implicated in the interaction with other proteins and may regulate the activity of the RNA-dependent RNA polymerase (RdRP) (Kirchdoerfer and Ward, 2019).

Overall, six amino acid mutations were identified in RdRP in the Russian sequences. These mutations most likely do not affect remdesivir binding to the RdRP protein of SARS-CoV-2 (Shannon et al., 2020). So, no naturally occurring resistant mutants were found.

The common mutations in spike protein S Asp614Gly and in Arg203Lys and Gly204Arg in nucleoprotein N observed in European and Russian sequences are of interest. Triple mutations in the nucleocapsid region G₂₈₈₁→A, G₂₈₈₂→A, and G₂₈₈₃→C (due to amino acid changes) leading to Arg203Lys and Gly204Arg (N) might affect antibody formation and cause a different serological profile in patients. The protein N is required for virion assembly and causes an immune response.

The Asp614Gly mutation was mapped on the three-dimensional structure of protein S (Supplementary Material Figure 1). Due to the fact that Asp614 of one protein S subunit is likely to interact with Thr859 of another subunit of the trimer, the Asp614Gly mutation might induce conformational changes in the spike subdomain 2 and the entire trimer. This, in turn, could lead to a change in the antigenicity of the spike protein and the spatial location of the furin cleavage site, which is located next to subdomain 2. The Asp614Gly mutation was recently characterized by Daniloski et al. (2020). Via site-directed mutagenesis, the authors showed that the SARS-CoV-2 variant with Gly614 in the spike protein showed eight-fold more effective transduction in multiple cell lines, including human lung epithelial cells. Moreover, this variant was more resistant to cleavage, proposing a possible mechanism for the increased transduction. This may play an important role in designing vaccine candidates for SARS-CoV-2.

The unique mutation identified in the virus EPI ISL 428,852 isolated during the study, C₃₃₇₃→A, leads to Asp1036Glu substitution in the nsp3 papain-like protease (PLP) domain, which possesses several important functions: proteolytic activity of N-terminal polyprotein and deubiquitinating activity. Some representatives of coronavirus papain-like proteases also participate in counteracting the innate immune response by removing ISG15 (interferon-stimulated gene 15, ubiquitin-like molecule) from proteins. The role of this mutation in the interplay of host defense has yet to be determined in future experiments (Chen et al., 2015).

Unique mutations identified in ORF6 – Pro57Leu (EPI ISL 428,906) and Met58Thr (EPI ISL 430,097) – are both in the region of ORF6 protein interacting with the NUP98–RAE1 complex. These interactions lead to the inhibition of nucleocytoplasmic transport and, therefore, disrupt interferon signaling (Frieman et al., 2007). Met58 is conservative among the viral proteins interacting with NUP98–RAE1 (SARS-CoV/ORF6, VSV/M, KSHV/ORF10) (Gordon et al., 2020). Thus, the mutations identified may disturb the functioning of the ORF6 protein.

A number of unique mutations have been identified in Russian SARS-CoV-2 variants. Their impact on the virus replicative cycle might be of interest.

Analysis of the molecular epidemiology of SARS-CoV-2 in Russia

Most of the Russian sequences clustered with the European sequences, except for four, corresponding to viruses isolated during March 20–24, 2020 in the city of Saint-Petersburg and in the Republic of Kabardino-Balkaria. Using a phylogenetic approach, the latter sequence was found to descend from the virus collected in Shanghai in late January. However, we did not find signs of direct

importation from China, most likely due to the lack of sequences from the Siberian and Far-Eastern parts of the country and sample collection times (Figure 3B).

Signs of internal circulation of imported viruses can be seen (Figure 3C), although the exact routes are difficult to establish due to the low mutation rate of the virus and comparatively small number of sequences. The sequences available from Saint-Petersburg allowed the estimation of the internal circulation start date, which was calculated to be during the first days of March. However, there are no earlier sequences in the database; thus it is difficult to estimate the correct time at which internal circulation started.

A more detailed epidemiological picture can only be achieved with more genetic data due to very low mutation rate of SARS-CoV-2.

Conclusions

Although the COVID-19 outbreak is still ongoing and investigations of cases are still in progress, we can conclude that most of the SARS-CoV-2 variants circulating in Russia contain a specific set of seven nucleotide mutations when compared to the original Wuhan virus, including amino acid substitutions in spike protein S and nucleoprotein N, possibly affecting their properties. Phylogenetic analysis of the sequences from Russia and other countries showed multiple importations of the virus, local circulation, and several patterns of virus spread across the territory of the country.

Funding source

This research was supported by grant of the Ministry of Science and Higher Education of the Russian Federation.

Ethical approval

The patients signed an informed consent form upon admission to the hospital. Additional approval was not required.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors would like to thank all of the physicians and personnel working with COVID-19 patients in GKB No. 40 DZM Hospital. The authors are grateful to all of the authors at the originating and submitting laboratories for sequences and metadata shared through GISAID. All submitters of the data can be contacted directly via www.gisaid.org. The authors would also like to thank Main Road Post LLC for time on computing clusters.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijid.2020.07.024>.

References

- Chen Y, Savinov SN, Mielech AM, Cao T, Baker SC, Mesecar AD. X-ray structural and functional studies of the three tandemly linked domains of non-structural protein 3 (nsp3) from murine hepatitis virus reveal conserved functions. *J Biol Chem* 2015;290(42):25293–306. doi:<http://dx.doi.org/10.1074/jbc.M115.662130>.
- Daniloski Z, Guo X, Sanjana NE. The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. , doi:<http://dx.doi.org/10.1101/2020.06.14.151357>.
- Decree of the President of the Russian Federation. Decree of the President of the Russian Federation of 02.04.2020 No. 239 "On measures to ensure sanitary and

- epidemiological welfare of the population in the Russian Federation in connection with the spread of a new coronavirus infection (COVID-19)". 2020. <http://kremlin.ru/events/president/news/63065>.
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20(5):533–4, doi:[http://dx.doi.org/10.1016/S1473-3099\(20\)30120-1](http://dx.doi.org/10.1016/S1473-3099(20)30120-1).
- Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P, Baric RS. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J Virol* 2007;81(18):9812–24.
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing [published online ahead of print, 2020 Apr 30]. *Nature* 2020;., doi:<http://dx.doi.org/10.1038/s41586-020-2286-9>.
- 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, doi:<http://dx.doi.org/10.1093/bioinformatics/bty407>.
- Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Comm* 2019;10(1):2342–28.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018;35(6):1547–9, doi:<http://dx.doi.org/10.1093/molbev/msy096>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome -coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clin Exp Pediatr*. 2020;63(4):119–24, doi:<http://dx.doi.org/10.3345/cep.2020.00493>.
- Pommié C, Levadoux S, Sabatier R, Lefranc G, Lefranc MP. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognit* 2004;17(1):17–32.
- Rabaan AA, Al-Ahmed SH, Haque S, Sah R, Tiwari R, Malik YS, et al. SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. *Infez Med* 2020;28(2):174–84.
- Sagunenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* 2018;4(1):vex042, doi:<http://dx.doi.org/10.1093/ve/vex042>.
- Shannon A, Le NT, Selisko B, Eydoux C, Alvarez K, Guillemot JC, et al. Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antiviral Res* 2020;178:104793.
- Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sciences. *Protein Sci* 2018;27:135–45.
- Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 2011;7:539, doi:<http://dx.doi.org/10.1038/msb.2011.75>.
- Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, et al. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. *Euro Surveill* 2020;25(13), doi:<http://dx.doi.org/10.2807/1560-7917.ES.2020.25.13.2000305>.
- WHO. WHO Director-General's opening remarks at the media briefing on COVID-19, 11 March 2020. World Health Organization; 2020 11 March 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020> (Accessed 5 May 2020).
- WHO. World Health Organization (WHO). Novel Coronavirus (2019-nCoV) Situation report – 1. Geneva: WHO; 2020 21 January 2020. Available from: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf> (Accessed 5 May 2020).
- Yin C. Genotyping coronavirus SARS-CoV-2 methods and implications. *Genomics* 2020;., doi:<http://dx.doi.org/10.1016/j.ygeno.2020.04.016> pii: S0888-7543(20)30318–30319.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020;382(8):727–33, doi:<http://dx.doi.org/10.1056/NEJMoa2001017>.
- COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). [accessed 10 July 2020]. <https://talk.ictvonline.org/taxonomy/> 2020 [accessed 5 May 2020]. <https://www.mos.ru/authority/documents/> 2020 [accessed 10 July 2020].