SPREAD OF VARIANTS WITH GENE N HOT SPOT MUTATIONS IN RUSSIAN SARS-COV-2 ISOLATES

Kiryanov SA ✉, Levina TA, Kirillov MYu

"DNA-Technology", Moscow, Russia

SARS-CoV-2 is a RNA coronavirus of the β-CoVs family responsible for the pandemic of the severe acute respiratory syndrome (COVID-19). It is hypothesized that the viral genome of the contemporary predominant founder clones is likely to be evolving in geographic-dependent manner. Thus, strains of different origin may be characterized by different mutation patterns. The study was aimed to perform the mutational and phylogenetic analysis of the Russian SARS-CoV-2 genomes at different time periods and in various regions, as well as to characterize the mutational profiles of isolates using the bioinformatics approaches. The mutation accumulation was compared in 86 SARS-CoV-2 whole-genome sequences from Russia and 220 from Europe and North America in order to reveal the characteristic gene variations, the possible positive selection patterns. Along with the known mutation variants in the structural proteins genes, typical for isolates of European origin, several additional mutations including the synonymous mutation in gene *M* (C26750T) characteristic for Russian isolates were revealed. Double mutation R203K and G204R in the nucleocapside gene which previously emerged in Europe began spreading and rapidly (within a month) became the dominant form in Russia. The results obtained indicate that the viral genome of most Russian isolates evolves with accumulation of new mutations associated with increased viral transmission. Data on the SARS-CoV-2 genome specific mutation patterns might be used for the detection of the virus, as well as for tracking and controlling of its spread.

Keywords: SARS-CoV-2, COVID-19, genome, coronavirus, mutations, Russian isolates

РАСПРОСТРАНЕНИЕ ВАРИАНТОВ С ЧАСТЫМИ МУТАЦИЯМИ В ГЕНЕ КАПСИДНОГО БЕЛКА N В РОССИЙСКИХ ИЗОЛЯТАХ SARS-COV-2

С. А. Кирьянов ✉, Т. А. Левина, М. Ю. Кириллов

ООО «ДНК-Технология», Москва, Россия

SARS-CoV-2 представляет собой РНК-вирус семейства β-коронавирусов, вызвавший пандемию тяжелого острого респираторного синдрома (COVID-19). Предполагают, что эволюционные изменения генома ныне преобладающих исходных клонов могут быть обусловлены географическими особенностями, поэтому для штаммов разного происхождения могут быть характерны различные паттерны мутаций. Целью исследования было провести мутационный и филогенетический анализы геномов SARS-CoV-2 российского происхождения в разные периоды и в разных регионах, а также охарактеризовать мутационные профили изолятов, используя подходы биоинформатики. Проведено сравнение накопления мутаций в 86 полногеномных последовательностях SARS-CoV-2 из России и 220 из Европы и Северной Америки для выявления характерных генных вариаций, возможных паттернов селективного отбора. Помимо известных мутаций в генах структурных белков, типичных для изолятов европейского происхождения, выявлены дополнительные синонимичные мутации российского происхождения, в том числе в гене *M* (C26750T). Двойная мутация R203K и G204R в гене нуклеокапсида, ранее появившаяся в Европе, распространилась и быстро (в течение месяца) стала доминирующим вариантом в России. Полученные результаты свидетельствуют о том, что геном значительной части изолятов SARS-CoV-2 из России эволюционирует с накоплением новых мутаций, способствующих повышению трансмиссивности вируса. Данные о специфичных мутационных паттернах генома SARS-CoV-2 могут быть использованы для выявления вируса, отслеживания и контроля его распространенности.

Ключевые слова: SARS-CoV-2, COVID-19, геном, коронавирус, мутации, российские изоляты

COVID-19 pandemic is characterized by rapid spread of the virus in many (over 187) countries of the world [1]. The prevalence, mortality and severity of the disease vary significantly between geographic regions, countries, and among the age groups of infected people [2–4].

The differences can be explained by geographic-dependent pattern of SARS-CoV-2 genome evolution and differentiation (due to quarantine measures and physical distancing), and by formation of several types from the ancestral Wuhan type (Hubei province, China) [5, 6]. RNA viruses appear to be characterized by high mutation rate. The emerging non-adaptive mutations lead to elimination of the strain from the population. The emerging adaptive mutations can be expected to provide the strain with a selective advantage usually in the form of high mutation rate and, consequently, of higher transmission rate.

The rapid spread of SARS-CoV-2 raises questions whether its evolution is driven by adaptive mutations and, if so, by which mutations in which genes.

The coronavirus genome is comprised of approximately 29,900 nucleotides. It encodes the extended open reading frame 1ab (ORF1ab) polyprotein, functioning as replicase and polymerase complex, and four structural proteins: membrane (M) protein, spike (S) glycoprotein, envelope (E) protein and nucleocapsid (N) phosphoprotein (the same as in all other β-CoVs) [7].

SARS-CoV-2 genome is being intensively studied aimed to diagnose the infection, to assess its pathogenic potential, and to track the evolution. To date, the GISAID database comprises over 25,000 viral genomic sequences collected in a few dozen countries. Previously, the mutation patterns of *ORF1ab* gene,

S glycoprotein gene and the non-structural proteins nsp6 and nsp8 genes were used for the SARS-CoV-2 genome evolution tracking [8–10]. Thus, the hot spot mutations C241T, C3037T and C14408T in the *ORF1ab* gene encoding the replication complex proteins were detected in the SARS-CoV-2 genomes isolated in Western Europe, together with the hot spot mutation A23403G in the S gene, the products of which interact with the ACE2 receptor. In Western European patients, the course of COVID-19 infection was more severe than in patients from other geographic regions [11]. The listed co-mutations combination determined the clade 20A (formerly known as clade G) is likely to be responsible for the enhanced transmission of the variant and for its being a dominant form in Europe.

The SARS-CoV-2 genome evolution record remains incomplete since the current reports cover mainly isolates retrieved in the USA, European countries, China, and some other countries. Particularly, there is a lack of data about SARS-CoV-2 genome mutational profiles retrieved in Russia. This study is of special relevance in connection with the observed contrast between the rapid viral expansion and low lethality rate in Russia.

The study was aimed to perform the mutational and phylogenetic analysis of the Russian SARS-CoV-2 genomes at different time periods and in various regions, as well as to characterize the mutational profiles of isolates using the bioinformatics approaches.

METHODS

From March 1 to April 29, 2020, subset of 86 SARS-CoV-2 nucleotide sequences isolated from Russian patients and 220 sequences collected in Europe and the USA were selected for analysis (were downloaded from the NCBI and GISAID databases). Inclusion criteria: full-length sequence of 26,000–30,000 nucleotides, sequences annotated as SARS-CoV-2. Exclusion criteria: re-submitted sequences, sequences containing too many undefined nucleotides. The multiple sequences alignment was performed using Clustal Omega (EMBL-EBI; Great Britain) and Blast (NCBI; USA). The MT233519 sequence, SARS-CoV-2/human/ESP/Valencia5/2020B, was considered a reference sequence for the analysis of isolates sampled in Russia.

Nextstrain (https://nextstrain.org/) was used for the SARS-CoV-2 sequences phylogenetic analysis, temporal dating of ancestral nodes, as well as for discrete traits, frequency and anchor mutations emergence dates reconstruction across the tree [12].

The B-cell epitopes were predicted based on the analysis results with the algorithm previously proposed for SARS-CoV by a group of researchers [13]. The following prediction tools were used for the primary N phosphoprotein amino acid sequence: BepiPred-2.0 (DTU; Denmark) for the linear B-cell epitopes prediction [14], and DiscoTope 2.0 (DTU; Denmark) for the conformational B-cell epitopes prediction [15]. Both tools were provided by the IEDB Immunobrowser resource (NIAID; USA).

When attempting to predict the linear B-cell epitopes with BepiPred-2.0 (DTU; Denmark) [14], the maximum threshold value was set at 0.75, the specificity was >0.85, and the sensitivity was <0.40. The sequences of more than 7 amino acid residues were analyzed. When predicting the conformational B-cell epitopes with DiscoTope 2.0 (DTU; Denmark) [15], the predictive positive value (PPV) was >–3.7, the specificity was ≥ 0.75, and the sensitivity was <0.40.

RESULTS

The total of 86 SARS-CoV-2 whole-genome sequences isolated from Russian patients in March–April 2020 was analyzed. Of those, 38 isolates (44%) were collected in March, and 56 (56%) were isolated by the end of April 2020. The SARS-CoV-2 genome nucleotide sequences were aligned and compared with those isolated from 220 Europeans and Americans (selected randomly from the GISAID global database). Phylogenetic analysis of the selected nucleotide sequences of Russian and European ancestry showed that all Russian isolates except one belonged to clade 20A (previously known as G) (Fig. 1). All other sequences except one sample from the Kabardino-Balkarian Republic identified in March carried the mutation A23403G with the substitution D614G in the S glycoprotein gene, and mutation C14408T with the substitution P314L in the gene encoding the ORF1b protein, along with the synonymous mutations C241T and C3037T. In Europe the listed mutations had been previously detected in the isolate of German ancestry (Germany/BavPat1/2020), and later in isolates from Italy retrieved in February. All the observed mutations are likely to define the stable haplotype currently dominant in European isolates, isolates from the East Coast of the USA, and in Russian isolates.
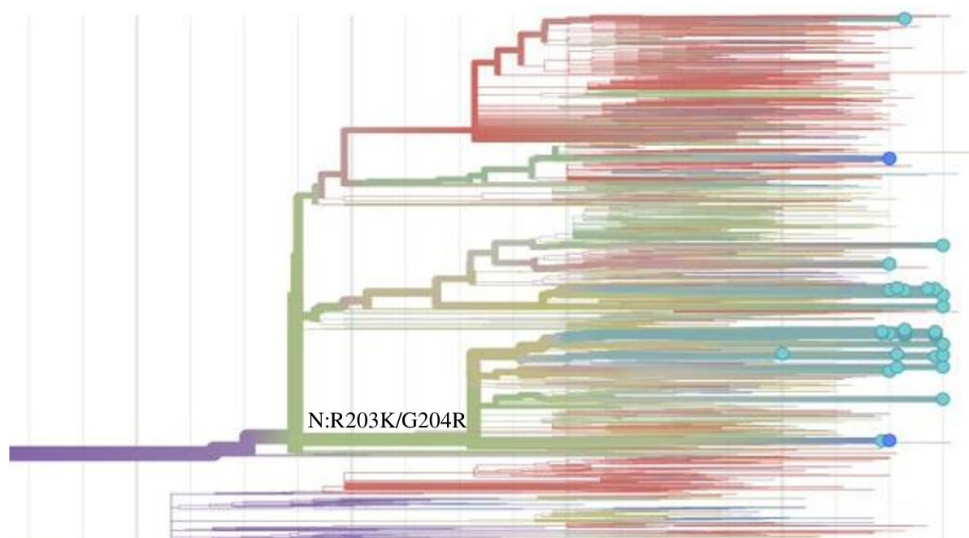

N:R203K/G204R

Fig. 1. Phylogenetic analysis of Russian, European, and American SARS-CoV-2 sequences, distinguished by the double mutation R203K and G204R in gene *N*. Russian isolates are highlighted in blue, European isolates in green and dark green, and American isolates in dark red

Table 1. Subclusters and characteristic mutations identified in the SARS-CoV-2 isolates from Russia

Subclusters: mutations in gene *N*	*ORF1ab* mutations	**Other genes mutations	***Additional mutations	Possible origin	Emergence time	Region of Russia and number of isolates
AP1 - R203K G204R		M: c26750t	ORF1a: P959S - 7 ORF1a: P309L 8	Italy Italy	02.03.2020 16.04.2020	Saint-Petersburg (40) Yakutia, Buryatia (5)
AP2 - R203K G204R	ORF1a: G3278S (g10097a) S: c23731t	ORF1a: A364S	ORF1b: M1499I	England	09–16.03.2020	Moscow (2) Yakutia (2)
		ORF1a: T1246I L3606F	ORF7a: L5F	England	02.03.2020	Moscow (4)
AP3 - R203K G204R		M: T175M	ORF1a: P892S, ORF1b: I1887V N: A152S	Italy	01–10.03.2020 14.03.2020	Lipetsk (2) Krasnodar (1) Moscow (3)
R203 G204	a20268g		N: N140T 8 N: N140K, T205I; ORF8: R101L	Spain	04–08.03.2020	Saint-Petersburg (14) Pskov (1) Omsk (1)
R203 G204	ORF1a: T265I (c1059t)	ORF3a: Q57H (g25563t)	ORF1a: P3395L S: G261V, ORF1a: L3606F c29200t	France Austria	04–16.03.2020 08–15.03.2020	Orenburg (2) Sakhalin (1) Chelyabinsk (1) Moscow (2)
R203 G204		M: D3G (a26530g)	G4255T ORF1a: L890F; ORF1b: Q348H, M1156I; ORF1b: L1701F, V1615I	England	04–05.03.2020 11.03.2020	Moscow (2) Moscow (2)

Note: * mutations in the *ORF1ab* gene defining the subgroup of the isolate; ** mutations in other genes defining the subgroup of the isolate; *** other frequent mutations identified in the isolates.

By the presence or lack of a triple mutation G28881A, G28882A, and G28883C in the *N* gene causing double nonsynonymous mutation R203K and G204R, genome sequences of SARS-CoV-2 isolates from Russia can be divided into two unequal groups of 59 and 26 sequences, respectively. Phylogenetic analysis of isolates retrieved in Russia, Europe and the USA reveals that the double mutation R203K and G204R previously discovered in the isolate from Valencia, Spain (MT233522, March 2, 2020) also forms a distinct subclade 20B (Fig.1). It should be noted that in contrast to Russian isolates the most European and American isolates form clades with lack of the triple mutation G28881A, G28882A and G28883C.

The subclade of Russian isolates defined by the double mutation R203K and G204R is subdivided into three unequal groups. The most numerous group is the group named AP1, comprising more than 40 isolates mainly from Saint-Petersburg and, apparently, of Italian origin. This group diverged from its predecessor is defined by the synonymous mutation C26750T in gene *M,* specific substitution only for these Russian isolates. The time of emergance of this mutation is not later than early March. The group is also characterized by the microclonality effect, defined by the accumulation of mostly synonymous mutations in the 5' region of the gene *ORF1ab*, and divided genome variants into additional subpopulations. Detailed information about defined subgroups and mutations is presented in Table 1.

The group AP2 comprising six isolates from Moscow and two from Yakutia is defined by the mutations in gene *ORF1a* (G3278S, T1246I, L3606F) and the synonymous mutation C23731T in gene *S*. The subgroup of four isolates from Moscow and Yakutia subsequently diverged from the ancestor with accumulation the mutation A364S and an additional mutation M1499I in the gene *ORF1ab*. The latter is exclusively found in isolates of Russian origin till now, it emerged before the middle of March.

Another group AP3 (six isolates from Moscow, Lipetsk and Krasnodar) probably originated from Italy has a characteristic mutation T175M in gene *M*. Further these isolates also differ

by mutations in gene *ORF1ab* (P892S, I1887V). The less numeroust group of three isolates (Moscow) is defined by the additional mutation A152S in gene N. This mutation is probably of Russian origin, and identified no later than the middle of March.

In the group of 26 isolates without mutation 203K and G204R in gene *N*, the accumulation of mutations also occurs mainly in the gene *ORF1ab*. The most common differentiating mutation is a synonymous mutation at the position A20268G of Spanish origin (found in 16 isolates, mostly from Saint-Petersburg). The presence of nonsynonymous mutations in the gene *ORF1ab* (T265I, P3395L, etc.), as well as in the genes *ORF3a* (Q57H) and *M* (D3G) allowed to identify several subgroups with insignificant the number of isolates (4–6). For information on the isolates' origin and additional mutations see Table 1. Only three isolates of 26 were found to carry additional nonsynonymous mutations in gene *N*: double mutation N140K and T205I, as well as N140T and A397V.

Thus, regardless of their origin, the SARS-CoV-2 variants with double mutation R203K and G204R in the gene N are a dominant form in various regions of Russia.

To determine the time of appearance and distribution double mutations R203K and G204R in Russia the analysis of most abundant viral genomes obtained from patients from Moscow and Saint-Petersburg in March–April 2020 classified according to the sampling date (available from GISAID) was performed. Four time-period subgroups according to the emergence date were identified as follows: March 10–12, 2020 (genomes collected from eight patients), March 19–21, 2020 (genomes from nine patients), April 1–3, 2020 (genomes from 16 patients), April 10–12, 2020 (genomes from 29 patients). The number of other accumulated mutations (mainly in the gene *ORF1ab*) changed during each time period: was 2, 4 and 4 in the group of genomes with double mutation R203K and G204R, and 2, 3 and 3 in the group of genomes with no the mutation. The latter group included two isolates with additional mutations in gene *N*: double mutation N140K and T205I, and N140T. Divergence of other genes did not affect the distribution of variants with double mutation R203K and G204R in gene *N*. In late March
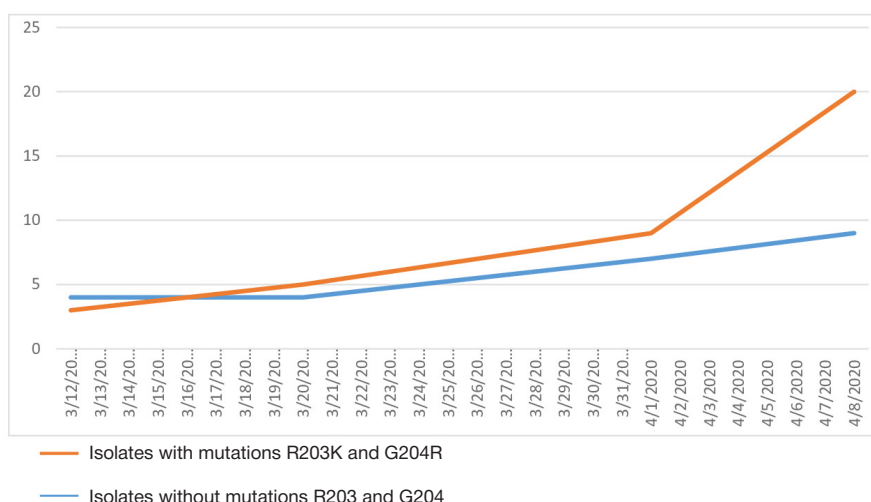
Fig. 2. Abundance of SARS-CoV-2 isolates carrying mutations R203K and G204R in gene *N* (Russia, March–April 2020)

and early April the proportion of isolates with double mutation R203K and G204R was more than doubled, by the middle of April it was more than 69.5% (Fig. 2).

The overall distribution and abundance mutation patterns in the *N* gene s in the nucleotide sequences of isolates from Europe and the United States deposited in the GISAID and NCBI databases was verified. It is worth noting that in European populations the abundance of subclade with double mutation R203K and G204R in gene *N* was significantly lower than in Russia, and presented in 32.6 % (1068 genomes out of 3241). In the USA the abundance of the same subclade was even lower (13.3%, 464 genomes out of 3479). The distribution of nonsynonymous mutations in gene *N* turned out to be uneven: 58.7% of mutations were located within the $N_{179-217}$ region. Using the linear B-cell epitope prediction tool, two possible linear B-cell epitope variants were predicted in the N protein at the positions of 23–36 and 178–207 respectively with the maximum threshold value at >0.758, (Table 2). Using the appropriate tool the conformational B-cell epitopes in the *N* gene with the threshold value >–0.37 and specificity 0.75 were predicted in about the same positions (26–36 and 193–207). The flanking positions R203, G204 and T205 within predicted B-cell epitope peptide were also noted (no data reported). However, about 25% of amino acid residues may be predicted as a part of the B-cell epitope incorrectly due to specificity of 0.75.

DISCUSSION

The data reported indicate that the SARS-CoV-2 genome is evolves forming several types clustered in distinct groups in geographic-dependent manner [16]. The mutation analysis of geographic-dependent isolates provides an insight into the hot spot mutation patterns responsible for high transmissibility of the virus. It has been reported, that at least five major mutations (C241T, C3037T, T28144C, C14408T, A23403G) turned out to be the most abundant in the Western European SARS-CoV-2 isolates [11]. The listed co-mutations, probably formed clade 20A, are likely to be responsible for the increased transmission of the virus and for its being a dominant form in Europe. According to mutational and phylogenetic analysis of SARS-CoV-2 genomes

isolated in Russia in March-April 2020, clade 20A appears to be one of the most widespread, which indicates European origin of Russian isolates. However, in Russia, unlike Western Europe, the subclade 20B characterized by the triple mutation G28881A, G28882A and G28883C which results in double substitution R203K and G204R in the N protein has spread and has become a dominant form. Thus, in Russia at the end of April the abundance of genomes with the double mutation R203K and G204R was over 69.5%, while in Europe it was 32.6 %. In the USA the number of genomes belonging to the same subclade defined by mutations R203K and G204R was even lower and accounted for 13.3%. The observed variant was likely to start circulating in Russia in early to mid-March 2020. The further expansion of the variant was accompanied by the formation of new subtypes with accumulation characteristic mutations in gene *M* (C26750T) or *ORF1b* (M1499I or G17964T), following subsequent divergence due to new single (mostly synonymous) mutations in the gene *ORF1ab*. Rapid spread of the variant with double mutation R203K and G204R in gene *N* may be indicative of its adaptability and ability to increase the transmission rate rather than modulate the virulence.

The functional effect of the mutant AAACGA motif in the nucleocapsid gene remains uncertain. The N protein appears to be responsible for the formation of the helical nucleocapsid during the virion assembly, and also plays a key role in replication and transcription. The protein is able to elicit the immune response and therefore may become a potential target for vaccine development [17]. The localization of potential B-cell and T-cell epitopes in the S glycoprotein, membrane M protein and capsid phosphoprotein N, predicted using the homologous regions of the SARS-CoV viral genome has been previously reported [18]. Our attempt at mapping the predicted $N_{179-207}$ B-cell epitope peptide amino acid sequence of the protein N allows suggesting that the positions R_{203} and G_{204} are located within the epitope. Mutations R203K and G204R result in two strong positively charged amino acid residues in close positions, in contrast to one positively charged residue in the wild type genotype, which may contribute to a decreased conformational entropy compared to the initial genotype. Currently, the bioinformatics methods without experimental data support do

Table 2. SARS-CoV2 B-cell epitope variants predicted in the nucleocapsid N protein

Epitope type	Position in the N protein	Mutation position	Epitope amino acid sequence
B-cell	178–207	R203K G204R T205I	GGSQASSRSSSRSRNSSRNSTPGSSRGTSP
B- cell	23–36		STGSNQNGERSGAR

not allow us to assess the biological significance of these mutations. Moreover, there is no reason to link the prevalence of these mutations of the SARS-CoV-2 in Russia with the viral pathogenicity. Further study of the SARS-CoV-2 viral genome evolution will allow the researchers not only to monitor the current epidemiological processes, but also optimize the existing RT-PCR diagnostic tests and search for new targets for vaccine development.

CONCLUSION

The current data indicate that the vast majority of SARS-CoV-2 isolates from Russia is of European origin. The viral genome of

the most Russian isolates evolves with the accumulation the new mutations associated with increased viral transmission. The double mutation R203K and G204R in the nucleocapsid gene has begun spreading and has rapidly become the dominant form in Russia.

Identification of the SARS-CoV-2 genome variants characteristic to the Russian population provides an insight into their further adaptive evolution. Data on the SARS-CoV-2 genome characteristic mutation patterns including the mutation patterns of the genes for structural proteins N and M might be used for the detection of the virus, as well as for tracking and controlling of its spread.

References

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579 (7798): 270–3. PubMed PMID: 32015507.
2. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020; 20 (6): 669–7. PubMed PMID: 32240634.
3. Spychalski P, Błażyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. Lancet Infect Dis. 2020; 20 (7): 774–5. PubMed PMID: 32243815.
4. Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of Covid-19/SARS-CoV-2. Osong Public Heal Res Perspect. 2020; 11 (2): 74–80. PubMed PMID: 32257772.
5. Phan T. Novel coronavirus: from discovery to clinical diagnostics. Infect Genet Evol. 2020; 79: 104211. PubMed PMID: 32007627.
6. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci. 2020; 117 (17): 9241–3. PubMed PMID: 32269081.
7. Boopathi S, Poma A, Kolandaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. J Biomol Struct Dyn. 2020: 1–10. PubMed PMID: 32306836.
8. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020; 18: 179. Available from: https://doi.org/10.1186/s12967-020-02344-6.
9. Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. J Infect. 2020; 81 (1): 24–7. PubMed PMID: 32283146.
10. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, et al. Whole genome and phylogenetic analysis of two SARSCoV-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): 1–5. PubMed PMID: 2000305.
11. Yin C. 2020 Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics. 2020; 112 (5): 3588-6. PubMed PMID: 32353474.
12. Hadfield J, Megill K, Bell SM. Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatica. 2018; 34 (23): 4121–3. PubMed PMID: 29790939.
13. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe. 2020; 27 (4): 671–80. PubMed PMID: 32183941.
14. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45 (1): 24–9. PubMed PMID: 28472356.
15. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol. 2012; 8 (12): e1002829: 1–10. PubMed PMID: 23300419.
16. Hahn G, Lee S, Weiss S. and Lange C. Unsupervised cluster analysis of SARS-CoV-2 genomes reflects its geographic progression and identifes distinct genetic subgroups of SARS-CoV-2 virus. bioRxiv. Preprint. 2020 Jun 30. Available from: https://www.biorxiv.org/content/10.1101/2020.05.05.079061v2
17. Zhao P, Cao J, Zhao L-J, Qin Z-L, Ke J-S, Pan W, Ren H, et al. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. Virology. 2005; 331 (1): 128–5. PubMed PMID: 15582659.
18. Koyama T, Weeraratne D, Snowdon J, and Parida L. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. Pathogens. 2020; 9 (5): 324. PubMed PMID: 32357545.

Литература

1. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579 (7798): 270–3. PubMed PMID: 32015507.
2. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020; 20 (6): 669–7. PubMed PMID: 32240634.
3. Spychalski P, Błażyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. Lancet Infect Dis. 2020; 20 (7): 774–5. PubMed PMID: 32243815.
4. Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of Covid-19/SARS-CoV-2. Osong Public Heal Res Perspect. 2020; 11 (2): 74–80. PubMed PMID: 32257772.
5. Phan T. Novel coronavirus: from discovery to clinical diagnostics.

Infect Genet Evol. 2020; 79: 104211. PubMed PMID: 32007627.
6. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci. 2020; 117 (17): 9241–3. PubMed PMID: 32269081.
7. Boopathi S, Poma A, Kolandaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. J Biomol Struct Dyn. 2020: 1–10. PubMed PMID: 32306836.
8. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020; 18: 179. Available from: https://doi.org/10.1186/s12967-020-02344-6.
9. Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation

of Non-Structural Protein 6 (NSP6) could affect viral autophagy. J Infect. 2020; 81 (1): 24–7. PubMed PMID: 32283146.

10. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, et al. Whole genome and phylogenetic analysis of two SARSCoV-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): 1–5. PubMed PMID: 2000305.

11. Yin C. 2020 Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics. 2020; 112 (5): 3588-6. PubMed PMID: 32353474.

12. Hadfield J, Megill K, Bell SM. Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatica. 2018; 34 (23): 4121–3. PubMed PMID: 29790939.

13. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe. 2020; 27 (4): 671–80. PubMed PMID: 32183941.

14. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45 (1): 24–9. PubMed PMID: 28472356.

15. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol. 2012; 8 (12): e1002829: 1–10. PubMed PMID: 23300419.

16. Hahn G, Lee S, Weiss S. and Lange C. Unsupervised cluster analysis of SARS-CoV-2 genomes reflects its geographic progression and identifes distinct genetic subgroups of SARS-CoV-2 virus. bioRxiv. Preprint. 2020 Jun 30. Available from: https://www.biorxiv.org/content/10.1101/2020.05.05.079061v2

17. Zhao P, Cao J, Zhao L-J, Qin Z-L, Ke J-S, Pan W, Ren H, et al. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. Virology. 2005; 331 (1): 128–5. PubMed PMID: 15582659.

18. Koyama T, Weeraratne D, Snowdon J, and Parida L. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. Pathogens. 2020; 9 (5): 324. PubMed PMID: 32357545.