Identification of the nucleotide substitutions in 62 SARS-CoV-2 sequences from Turkey

AYŞE BANU DEMİR
DOMENICO BENVENUTO
YUSUF HAKAN ABACİOĞLU
SILVIA ANGELETTI
MASSIMO CICCOZZI

Follow this and additional works at: https://journals.tubitak.gov.tr/biology

Recommended Citation
DEMİR, AYŞE BANU; BENVENUTO, DOMENICO; ABACİOĞLU, YUSUF HAKAN; ANGELETTI, SILVIA; and CICCOZZI, MASSIMO (2020) "Identification of the nucleotide substitutions in 62 SARS-CoV-2 sequences from Turkey," Turkish Journal of Biology. Vol. 44: No. 7, Article 6. https://doi.org/10.3906/biy-2005-69
Available at: https://journals.tubitak.gov.tr/biology/vol44/iss7/6

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Biology by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.
Identification of the nucleotide substitutions in 62 SARS-CoV-2 sequences from Turkey

Ayşe Banu DEMİR1*, Domenico BENVENUTO2, Hakan ABACIOĞLU3, Silvia ANGELETTI4, Massimo CICCOZZI2

1Department of Medical Biology, Faculty of Medicine, İzmir University of Economics, İzmir, Turkey
2Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy
3Department of Medical Microbiology, Faculty of Medicine, İzmir University of Economics, İzmir, Turkey
4Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome, Italy

* Correspondence: banu.demir@ieu.edu.tr

Abstract: A previously unknown coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been shown to cause coronavirus disease 2019 (COVID-19) pandemic. The first case of COVID-19 in Turkey has been declared in March 11th, 2020 and from there on, more than 150,000 people in the country have been diagnosed with the disease. In this study, 62 viral sequences from Turkey, which have been uploaded to GISAID database, were analyzed by means of their nucleotide substitutions in comparison to the reference SARS-CoV-2 genome from Wuhan. Our results indicate that the viral isolates from Turkey harbor some common mutations with the viral strains from Europe, Oceania, North America and Asia. When the mutations were evaluated, C3037T, C14408T and A23403G were found to be the most common nucleotide substitutions among the viral isolates in Turkey, which are mostly seen as linked mutations and are part of a haplotype observed high in Europe.

Key words: SARS-CoV-2, evolution, mutation, COVID-19

1. Introduction
Coronaviruses (CoV) are enveloped positive-stranded RNA viruses that belong to the family Coronaviridae and are divided into 4 genera which are alpha-CoV, beta-CoV, gamma-CoV, and delta-CoV. Similar to severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), SARS-CoV-2 is a beta-coronavirus (Gorbalenya et al., 2020). It encodes several structural proteins including envelope (E), membrane (M), nucleocapsid (N) and spike (S) proteins as well as non structural ones (Gorbalenya et al., 2020). The virus was shown to have gone through certain mutations both in its structural and non structural proteins within several months while spreading throughout the world (Pachetti et al., 2020; Wang et al., 2020).

Starting from December 2019, SARS-CoV-2 led to a worldwide COVID-19 pandemic, which caused more than 3 million cases along with more than 250,000 deaths within 5 months1. The first case of COVID-19 in Turkey was announced in March 11th and as of May 24th, the number of positive cases and deaths reached to 156,827 and 4,340, respectively2. A total of 63 sequences from SARS-CoV-2 isolates of Turkey were uploaded to global initiative on sharing all influenza data (GISAID) database between the dates March 25th and May 22nd3.

The aim of this study is to reveal the most common mutations of SARS-CoV-2 viral isolates from Turkey in comparison to the reference sequence from China (NC_045512.1). Our results revealed that some of the viral mutations are present in more than 60% of the isolates. Although further analysis and characterizations are needed, the data in this study may contribute to understanding the molecular evolution of SARS-CoV-2 in Turkey.

2. Materials and methods
2.1. Dataset construction
All the SARS-CoV-2 whole genome sequences have been downloaded from GISAID database4 (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017). The whole T.C. Sağlık Bakanlığı (2020). Türkiye’deki Güncel Durum [online]. Website https://covid19.saglik.gov.tr/ [accessed 05 May 2020].

Received: 18.05.2020  ●  Accepted/Published Online: 07.06.2020  ●  Final Version: 21.06.2020

This work is licensed under a Creative Commons Attribution 4.0 International License.
genome sequence dataset was constructed as including 63 viral sequences from Turkish patients that were submitted to the database between March 25th and May 22nd, 2020 (Supplementary Table 1) and the reference SARS-CoV-2 sequence isolated in Wuhan which was downloaded from GeneBank (NC_045512.1). One of the sequences (EPI_ISL_435057) was excluded from the analysis due to harboring extreme number of unique mutations, which can result from sequencing errors.

2.2. Nucleotide substitution analysis

SARS-CoV-2 isolate sequences from Turkey were compared to the reference SARS-CoV-2 sequence (NC_045512.1), by means of nucleotide substitutions. The constructed dataset was MAFFT\(^5\) aligned and manually edited using the AliView program to verify that the sequences were in frame. The nucleotide numbers were indicated starting from the 5' UTR of the viral sequence, while amino acid positions were indicated separately for each corresponding protein coding region\(^4\). The positions of the nucleotides and amino acids were further confirmed from GeneBank reference sequences (NC_045512.1)\(^4\). The nonconserved nucleotide positions were determined and the nucleotide substitutions were evaluated for their effects on amino acid changes by using the AliView and MEGA software.

3. Results

The age interval of the patients from Turkey, from whom the viral isolates were taken, was between 19 and 82. 41% of the patients were between the age range of 41–60, while 29% was between 19 and 40 and 28% was between 61 and 80 (Figure 1a) and the sex distribution was approximately equal (Figure 1b). Most of the uploaded samples seem to come from middle-west side of the country (Figure 1c).

When the SARS-CoV-2 strains from Turkey, compared to the reference viral sequence (NC 045512.1), some missense and silent mutations were identified. In more than 40% of the viral isolates, one or more nucleotide substitutions of C3037T, C14408T, A23403G, and G25563T were observed, which are present in the coding regions for the Nsp3, RNA-dependent RNA polymerase, spike glycoprotein and ORF3a protein, respectively (Table 1). Among these mostly seen substitutions, the ones seen in Nsp3, RNA-dependent RNA polymerase and spike glycoprotein were present in 61% (38/62) of the isolates from Turkey. Other viral genome regions\(^4\)NCBI (2020). GeneBank Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome [online]. Website https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2 [accessed 24 April 2020],\(^5\)MAFFT (2020). Multiple alignment program for amino acid or nucleotide sequences [online]. Website https://mafft.cbrc.jp/alignment/software/ [accessed 25 April 2020]. where the nucleotide substitutions were observed in 5 or more samples out of 62 include, Nsp6 (24/62), 3' to 5' exonuclease (22/62), Nsp2 (22/62), nucleocapsid protein (21/62), membrane glycoprotein (13/62), Nsp4 (11/62), Nsp3 (6/62) and Helicase (5/62) (Table 1).

Each of the 62 viral isolates were also evaluated by means of the mutations they harbor and 24 viral isolates found to have unique mutations in addition to the common mutations they harbor, that are not seen in other isolates from Turkey (Table 2; Supplementary Table 2). Most of the viral isolates found to have different nucleotide substitution combinations. However, the same nucleotide substitutions were observed for the samples EPI_ISL42874 and EPI_ISL_429871. Similarly, EPI_ISL437411 and EPI_ISL437413 were also found to have the same nucleotide substitution combination among them (Supplementary Table 2). The analyzed viral isolates were found to harbor 4 to 12 mutations per isolate compared to the reference sequence (Supplementary Table 2). C > T mutations also observed to predominate among the analyzed viral isolates.

4. Discussion

The first COVID-19 case in Turkey was declared in March 11th, almost after two and a half months from the first case declaration in China. When the nucleotide substitutions for the SARS-CoV-2 isolates in Turkey were analyzed compared to the reference genome of the virus from China, it was seen that during this time period, the virus had undergone several nucleotide substitutions, including silent and missense mutations.

When we consider the mutations in ORF1ab, nucleotide substitutions of C884T, G1397A, C3037T, G8653T, G11083T, C14408T, and C18877T were seen in more than 50% (31/62) of the isolates from Turkey. Other amino acids seem to have similar isoelectric points and this mutation is mostly seen in isolates from Europe, followed by North America (Pachetti et al., 2020). The mutation was found to be present in European isolates after February 20th, 2020 and thought to be associated with increased number of point mutations compared to isolates from Asia, which proposed to be somehow due to the presence of RNA polymerase within the proofreading machinery of the virus (Pachetti et al., 2020). A recent study indicates that SARS-CoV-2 genomes which harbor C14408T mutation, are more likely to have mutations in the membrane (M) and envelope (E) proteins (Eskier et al., 2020). Furthermore, recently revealed structure of the
replicating RNA polymerase of SARS-CoV-2 may further help to understanding of the effect of certain mutations within this protein (Hillen et al., 2020).

G11083T, corresponding to the amino acid substitution L37F within Nsp6 protein, was present in 38% of the samples (24/62) and this mutation was previously seen in SARS-CoV-2 sequences analyzed from all over the world (Benvenuto et al., 2020; Wang et al., 2020). In the study of Wang et al. (2020), C8782T substitution, which is also a silent mutation, was present in 28 out of 95 samples, although this mutation was only present in 2 samples (EPI_ISL_428718 and EPI_ISL_437317) in our study, which indicates that this mutation is not as common as in Europe for the viral isolates in Turkey and may be related with isolates from other countries. Both C8782T and G11083T mutations were found to be mostly present in Oceania isolates, where followed by North America and Europe subsequently (Pachetti et al., 2020). G1397A substitution in Nsp2 encoding region of ORF1a, which was present in 33% (21/62) of the isolates, was mainly seen in viral isolates from Oceania, however was also present in minor amounts in isolates from Asia and North America (Pachetti et al., 2020). This substitution leads to an amino acid change from valine to isoleucine at the position 198 (V198I) within Nsp2 protein, where both amino acids have the same isoelectric points.

A23403G mutation in the spike glycoprotein coding region was also amongst the mostly seen mutations in viral isolates from Turkey (61%). Spike glycoprotein functions to bind target receptor and facilitate membrane fusion and viral entry (Ou et al., 2020). This protein has 2 subunits, S1 and S2, where the former mediates attachment and the later mediates membrane fusion. A23403G substitution was found to be present in isolates from Europe and leads to an amino acid change from aspartate to glycine at position 614 (D614G) within the spike glycoprotein, where these amino acids differ by means of their isoelectric points (Pachetti et al., 2020). Another mutation found within the spike protein encoding region was C22444T, which is a silent mutation and seen in 7 out of 62 isolates (Table 1).

Similar to SARS-CoV, receptor binding domain (RBD) within the spike glycoprotein of SARS-CoV-2 seems to play a major role in viral infection by acting as an interaction point with target receptors on the host cell surface (Raj et al., 2013). In a recent study which performed multivariate generalized linear model (GLM) analysis with outpatient and hospitalized patients in the Sheffield Teaching Hospitals NHS Foundation Trust as the outcome revealed that patients carrying G614 mutation had higher viral loads compared to D614, although D614G status did not significantly affect the hospitalization status (Korber et al., 2020). It also seems that the viral isolates which carry
D614G mutation increases in number across the world and this mutation was proposed to have effect on the viral infectivity either due to its presence on the spike protein promoter surface region which might affect hydrogen bonding properties with neighbouring promoter regions or due to be in a site surrounded by antibody-dependent enhancement targets, where antibody binding may lead to a confirmational change that might increase the ACE2 interaction. Both mechanisms were proposed to play role in a more transmissible form of the virus (Korber et al., 2020). On the other hand, another study performed on 15,000 SARS-CoV-2 genomes indicated that the recurrent mutations do not increase transmissibility (Dorp et al., 2020). Therefore, the effect of D614G mutation on the transmission of the virus is still a debate.

Five out of 27 commonly seen mutations in viral isolates from Turkey were within the nucleocapsid phosphoprotein. One interesting finding was the presence

| Nucleotide substitution at the given position | Corresponding viral gene | Corresponding viral gene product | Amino acid change within the corresponding protein (if exists) | Mutation type | Number of samples seen (among 62 samples) | Percentage among 62 samples |
|---------------------------------------------|--------------------------|---------------------------------|-------------------------------------------------------------|---------------|----------------------------------------|-----------------------------|
| C3037T                                      | ORF1a                    | Nsp3                            | 106 (F)                                                     | Silent        | 38                                      | 61%                         |
| C14408T                                     | ORF1b                    | RNA-dependent RNA polymerase     | P323L                                                       | Missense      | 38                                      | 61%                         |
| A23403G                                     | S                        | Spike glycoprotein               | D614G                                                       | Missense      | 38                                      | 61%                         |
| G25563T                                     | ORF3a                    | ORF3a protein                    | Q57H                                                        | Missense      | 25                                      | 40%                         |
| G11083T                                     | ORF1a                    | Nsp6                            | L37F                                                        | Silent        | 24                                      | 38%                         |
| C18877T                                     | ORF1b                    | 3’ to 5’ exonuclease             | 280 (L)                                                     | Silent        | 22                                      | 35%                         |
| G29742T                                     | 3’ UTR                   |                                 |                                                             |               | 22                                      | 35%                         |
| G1397A                                      | ORF1a                    | Nsp2                            | V1981                                                       | Missense      | 21                                      | 33%                         |
| T28688C                                     | N                        | Nucleocapsid phosphoprotein      | 139(L)                                                      | Silent        | 21                                      | 33%                         |
| C241T                                       | 5’ UTR                   |                                 |                                                             |               | 20                                      | 32%                         |
| C26735T                                     | M                        | Membrane glycoprotein            | 71(Y)                                                       | Silent        | 13                                      | 20%                         |
| C26549T                                     | M                        | Membrane glycoprotein            | 9(T)                                                        | Silent        | 12                                      | 19%                         |
| C884T                                       | ORF1a                    | Nsp2                            | R27C                                                        | Missense      | 11                                      | 17%                         |
| G8653T                                      | ORF1a                    | Nsp4                            | M33I                                                        | Missense      | 11                                      | 17%                         |
| G28881A                                     | N                        | Nucleocapsid phosphoprotein      | R203K                                                       | Missense      | 9                                       | 14%                         |
| G28882A                                     | N                        | Nucleocapsid phosphoprotein      | G204R                                                       | Missense      | 9                                       | 14%                         |
| C228T                                       | 5’ UTR                   |                                 |                                                             |               | 8                                       | 13%                         |
| A9514G                                      | ORF1a                    | Nsp4                            | 320(L)                                                      | Silent        | 7                                       | 11%                         |
| C22444T                                     | S                        | Spike glycoprotein               | 294(D)                                                      | Silent        | 7                                       | 11%                         |
| G26720C                                     | M                        | Membrane glycoprotein            | 66(V)                                                       | Silent        | 7                                       | 11%                         |
| C28854T                                     | N                        | Nucleocapsid phosphoprotein      | S194L                                                       | Missense      | 7                                       | 11%                         |
| C5736T                                      | ORF1a                    | Nsp3                            | A1006V                                                      | Missense      | 6                                       | 10%                         |
| G9479T                                      | ORF1a                    | Nsp4                            | G309C                                                       | Missense      | 6                                       | 10%                         |
| T28835C                                     | N                        | Nucleocapsid phosphoprotein      | S188P                                                       | Missense      | 6                                       | 10%                         |
| C7765T                                      | ORF1a                    | Nsp3                            | 1682(S)                                                     | Silent        | 5                                       | 8%                          |
| C17690T                                     | ORF1b                    | Helicase                        | S485L                                                       | Missense      | 5                                       | 8%                          |
| T26551C                                     | M                        | Membrane glycoprotein            | V10A                                                        | Missense      | 5                                       | 8%                          |
of 2 subsequent missense mutations that are seen as a cluster in 9 out of 62 samples (14 %). These mutations were due to nucleotide substitutions in 3 nucleotides in order where 2 of them (G28881A and G28882A) results in arginine to lysine (R203K) substitution and the third one (G28883C) results in glycine to arginine (G204R) in the nucleocapsid phosphoprotein, where both substituted amino acids differ from their original amino acids in means of their isoelectric points (Pachetti et al., 2020) (Table 1). G28881A, along with A23403G substitution in spike glycoprotein, seems to have occurred after February 16th, 2020 in Europe (Pachetti et al., 2020). C28144T substitution in the nucleocapsid protein coding region, which leads to an amino acid change (Leu to Ser) in addition to B-type substitutions. A- and C-types are said to present mainly outside of East Asia, where B-type is said to be present mainly in East Asia. The isolates from Turkey analyzed in this study mainly harbor cytosine in nucleotide position 29095. Another study, which analyzed 622 complete SARS-CoV-2 genomes by an unrooted maximum likelihood tree divided the viral genomes into 3 clusters, which was mainly similar to the 3 viral variants identified in 162 SARS-CoV-2 genomes (Forster et al., 2020), and performed linkage analysis between the mutations seen within these clusters (Bai et al., 2020). According to the linkage analysis, C241T, C3037T, and G26144T synonymous mutation were seen in all 3 clusters also.

The present study shows the presence of the mutation in several strains all over the world including some isolates from Turkey. In addition to S protein, nucleocapsid protein was also proposed to be important in COVID-19 infectivity (Goh et al, 2020). Further studies are needed to clarify if the missense mutations within this region can be important in the infection strategy of the SARS-CoV-2 virus or not.

Table 2. Nucleotide substitutions present only in a single isolate among the analyzed viral isolates from Turkey.

| GISAID ID    | Nucleotide substitutions present only in the corresponding isolate |
|--------------|---------------------------------------------------------------------|
| EPI_ISL_424366 | G23876A, C29563T                                                   |
| EPI_ISL_427391 | C2997T                                                             |
| EPI_ISL_428368 | C12809T                                                            |
| EPI_ISL_428717 | C21304A, G21305A, C28054T                                           |
| EPI_ISL_428718 | C8782T, G14122T, G28878A                                           |
| EPI_ISL_428720 | G12248T; T23559A                                                   |
| EPI_ISL_428713 | C4524T                                                             |
| EPI_ISL429870  | C19170T; C25275T                                                   |
| EPI_ISL_429873 | C1437T                                                             |
| EPI_ISL_429864 | G944A                                                              |
| EPI_ISL429865  | C7834T; C26340T                                                   |
| EPI_ISL_429868 | C11074T                                                            |
| EPI_ISL_437306 | C8683A                                                             |
| EPI_ISL_437307 | T6202A, C8964T, C10202T, C16247T, C24865T                            |
| EPI_ISL_437308 | C15240T                                                            |
| EPI_ISL_437309 | C16616T, A23734T                                                   |
| EPI_ISL_437317 | G22468T, G25314T, T28144C                                           |
| EPI_ISL_437318 | C5477T, C6402T                                                     |
| EPI_ISL_437319 | G19285A                                                            |
| EPI_ISL_437328 | C1825T                                                             |
| EPI_ISL_437330 | C5826A                                                             |
| EPI_ISL_437331 | C12700T                                                            |
| EPI_ISL_437333 | T15102C                                                            |
| EPI_ISL_437335 | A27354G                                                            |
C14408T and A23403G in Cluster 3 were in complete linkage and the TTTG haplotype was high in Europe and correlated with the death rate. In the viral isolates analyzed in this study, C3037T, C14408T and A23403G, which were the most common mutations (61%), exist together. In 11 out of 62 samples, C241T was also observed to be present together with C3037T, C14408T and A23403G. The reason of not observing C241T in linkage with other mutations with the same percentage can be due to the absence of the first 265 nucleotides in 25 of the uploaded sequences to GISAID. However, there are also isolates that harbor either C241T and not the other 3 mutations or vice versa. This haplotype was proposed to be related with the high death rates in Europe (Bai et al., 2020). Analyzing the course of the COVID-19 disease in patients from whom the viral isolates were taken can give further information about the relatedness of this haplotype with the death rates in Turkey.

Apart from the common mutations, when we consider mutations seen in a single sample among analyzed isolates, some are not mentioned previously in the literature. C8782T was previously seen in more than 10 isolates in Guangdong province of China (Lu et al., 2020) and proposed to be clade specific in a study performed on 313 SARS-CoV-2 genomes (Li, Li, Cui, and Wu, 2020). G28878A, which is present in the same isolate with C8782T, was observed in isolates from Australia and USA (Li et al., 2020). Some of the observed mutations can be either unique to corresponding isolate or can also be a result of homoplasy or sequencing artefact since in an ongoing study, some sites within the viral genome are suspected to be homoplasic substitutions or sequencing artefacts6. G11083T is the most common one among such sites across different countries and sequencing technologies, which might be an indicator of this position being either a site for frequent mutation or an artefact. However, 38% of the samples (24/62) analyzed in this study harbor this mutation, although being sequenced by different technologies, which is consistent with this site being a site for frequent mutation.

Some homoplasic sites were found to be specific to certain sequencing technologies, such as the nucleotide position 11074. Nucleotide 3037 and 11074 were reported to be either artefacts or hypermutable low-fitness sites. However, 3037 was found to have a linkage with 3 other mutations and mainly observed in Europe. It is also among the mostly seen mutations along with 14408 and 23403 in the viral isolates analyzed. Therefore, it can be 1 of the hypermutable sites within the viral genome. Apart from 11074 and 3037, detailed analysis of the identified unique mutations by means of possible sequencing artefacts and homoplasic sites can reveal more information about them. Therefore, considering the possible sequencing artefacts while analyzing the sequences for substitution can be important (Korber et al., 2020).

Recombination is known to take place in evolution of coronaviruses and some breakpoints for recombination in SARS-CoV-2 was also reported (Korber et al., 2020; Rehman et al., 2020). Therefore, apart from single nucleotide mutations, identification of possible recombinational events can be important in vaccine development strategies.

In our dataset, 2 samples found to harbor quite lots of mutations, which both are sequenced with the same sequencing technology. Therefore, only the common mutations with other isolates in one of them were considered and the other one was excluded totally since the mutations it harbors were quite extreme. The number of extreme mutations might be due to the use of separate sequencing technology in these strains compared to the other isolates that are analyzed in this study.

The nucleotide substitutions showed that viral isolates from Turkey are genetically close to the ones from Europe, Middle East, North America and Asia. C3037T, C14408T and A23403G substitutions, which are present in Nsp3, RNA-dependent RNA polymerase and spike encoding regions respectively, were found to be the mostly seen mutations in Turkey SARS-CoV-2 isolates. Considering the missense mutations encountered in these isolates, further studies are needed how the identified amino acid changes affect the structure of the related proteins as well as the infectivity and spread of the virus. Also, the silent mutations within SARS-CoV-2 genome can be followed up to determine if any further missense mutations will take place within these regions, which may be helpful to understand the evolutionary strategy of the virus as it continues to evolve during its spread through the world.

Acknowledgments/Disclaimers
The authors acknowledge to all the researchers in originating and submitting labs who have shared the SARS-CoV-2 genome data on GISAID (http://www.gisaid.org/). The extended acknowledgement can be found as a supplementary file. No funding was used to conduct this research.

Conflict of interest
The authors declare no competing interests.
References

Bai Y, Jiang D, Lon JR, Chen X, Hu M et al. (2020). Evolution and molecular characteristics of SARS-CoV-2 genome. BioRxiv (preprint). doi: 10.1101/2020.04.24.058933

Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S et al. (2020). Evolutionary analysis of SARS-CoV-2: how mutation of non-structural protein 6 (NSP6) could affect viral autophagy. Journal of Infection (in press). doi: 10.1016/j.jinf.2020.03.058

Dorp LV, Richard, Shaw LP (2020). No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. BioRxiv (preprint). doi: 10.1101/2020.05.21.108506

Eden JS, Rockett R, Carter I, Rahman H, De Ligt J et al. (2020). An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. Virus Evolution 6 (1): 1-4. doi: 10.1093/ve/veaa027

Elbe S, Buckland-Merrett G (2017). Data, disease and diplomacy: GISAID’s innovative contribution to global health. Global Challenges 1 (1): 33-46. doi: 10.1002/gch2.1018

Elbe S, Buckland-Merrett G, Suner A, Oktay Y (2020). RdRp mutations are associated with SARS-CoV-2 genome evolution. BioRxiv (preprint). doi: 10.1101/2020.05.20.104885

Forster P, Forster L, Renfrew C, Forster M (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. Proceedings of the National Academy of Sciences 117 (17): 9241-9243. doi: 10.1073/pnas.2004999117

Goh GKM, Keith Dunker A, Foster JA, Uversky VN (2020). Rigidity of the outer shell predicted by a protein intrinsic disorder model sheds light on the COVID-19 (Wuhan-2019-nCoV) infectivity. Biomolecules 10 (2): 2019-2021. doi: 10.3390/biom10020331

Gorbalenya AE, Baker SC, Baric RS, De Groot RJ, Drosten C et al. (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology 5 (4): 536-544. doi: 10.1038/s41564-020-0695-z

Hillen HS, Kocic G, Farnung L, Dienemann C, Tegunov D et al. (2020). Structure of replicating SARS-CoV-2 polymerase. BioRxiv (preprint). doi: 10.1038/s41586-020-2368-8

Korber B, Fischer W, Gnanakaran SG, Yoon H, Theiler J et al. (2020). Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. BioRxiv (preprint). doi: 10.1101/2020.04.29.069054

Li J, Li Z, Cui X, Wu C (2020). Bayesian phylodynamic inference on the temporal evolution and global transmission of SARS-CoV-2. Journal of Infection: 1-12 [Epub ahead of print]. doi: 10.1016/j.jinf.2020.04.016

Lu J, Plessis L du, Liu Z, Hill V, Kang M et al. (2020). Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell 181 (5): 997-1003.e9. doi: 10.1016/j.cell.2020.04.023

Ou X, Liu Y, Lei X, Li P, Mi D et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nature Communications 11 (1): 1620. doi: 10.1038/s41467-020-15562-9

Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E et al. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent RNA polymerase variant. Journal of Translational Medicine 18 (179): 1-9. doi: 10.1186/s12967-020-02344-6

Raj VS, Mou H, Smits SL, Dekkers DHW, Müller MA et al. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495 (7440): 251-254. doi: 10.1038/nature12005

Rehman SU, Shafique L, Ihsan A, Liu Q (2020). Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. Pathogens 9 (3): 1-12. doi: 10.3390/pathogens9030240

Shu Y, McCauley J (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 22 (13): 2-4. doi: 10.2807/1560-7917.ES.2017.22.13.30494

Wang C, Liu Z, Chen Z, Huang X, Xu M et al. (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. Journal of Medical Virology 92 (6): 667-674. doi: 10.1002/jmv.25762
**Supplementary Table 1.** General information taken from GISAID about the SARS-CoV-2 sequences used in this study. (NA: Not available)

| GISAID ID     | Location | Sex   | Age | Collection date | Submission date | Travel history/other                      | Sequencing technology |
|---------------|----------|-------|-----|-----------------|-----------------|------------------------------------------|----------------------|
| EPI_ISL_429873| Kocaeli   | Male  | 71  | 2020-03-23      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429872| Kocaeli   | Female| 50  | 2020-03-25      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429871| Ankara    | Male  | 77  | 2020-03-23      | 2020-04-24      | Patient travelled to Saudi Arabia        | Illumina MiSeq       |
| EPI_ISL_429870| Sakarya   | Female| 57  | 2020-03-22      | 2020-04024      | Patient travelled to Saudi Arabia        | Illumina MiSeq       |
| EPI_ISL_429869| Konya     | Female| 59  | 2020-03-17      | 2020-04-24      | Patient travelled to Saudi Arabia        | Illumina MiSeq       |
| EPI_ISL_429868| Eskişehir| Female| 79  | 2020-03-17      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429867| Balıkesir | Female| 72  | 2020-03-17      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429866| Afyon     | Female| 52  | 2020-03-16      | 2020-04-24      | Patient travelled to Saudi Arabia        | Illumina MiSeq       |
| EPI_ISL_429865| Çanakkale | Female| 72  | 2020-03-18      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429864| Sakarya   | Male  | 33  | 2020-03-22      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429863| Sakarya   | Female| 42  | 2020-03-22      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_429862| Ankara    | Male  | 65  | 2020-03-22      | 2020-04-24      | Patient travelled to Saudi Arabia        | Illumina MiSeq       |
| EPI_ISL_429861| Ankara    | Male  | 48  | 2020-03-22      | 2020-04-24      | NA                                      | Illumina MiSeq       |
| EPI_ISL_428723| Aksaray   | Male  | 48  | 2020-03-22      | 2020-04-21      | Patient travelled to Saudi Arabia        | Illumina MiSeq       |
| ID       | Location | Gender | Age | Date Received | Date Sequenced | Notes                                      | Assembly Details                                                                 |
|----------|----------|--------|-----|---------------|----------------|--------------------------------------------|-----------------------------------------------------------------------------------|
| EPI_ISL_428722 | Balıkesir | Female | 37  | 2020-03-22    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428721 | Ankara   | Male   | NA  | 2020-03-21    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428720 | Ankara   | Female | 35  | 2020-03-21    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428719 | Siirt     | Male   | 52  | 2020-03-21    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428718 | Kocaeli   | Male   | 35  | 2020-03-19    | 2020-04-21     | Patient travelled to Saudi Arabia           | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428717 | Kocaeli   | Male   | 38  | 2020-03-19    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428716 | Ankara   | Female | 62  | 2020-03-18    | 2020-04-21     | Patient travelled to Saudi Arabia           | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428715 | Nevşehir  | Female | 55  | 2020-03-18    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428714 | Kastamonu | Male   | 60  | 2020-03-18    | 2020-04-21     | Patient travelled to Saudi Arabia           | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428713 | Ankara   | Female | NA  | 2020-03-18    | 2020-04-21     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428712 | Karaman  | Male   | 72  | 2020-03-17    | 2020-04-21     | Patient travelled to France                 | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 1,000x coverage     |
| EPI_ISL_428368 | İstanbul | Female | 49  | 2020-04-16    | 2020-04-20     |                                             | Illumina NextSeq Assembly: BWA-MEM 1,750x coverage                               |
| EPI_ISL_428346 | İstanbul | Male   | 49  | 2020-04-17    | 2020-04-20     |                                             | Illumina Next Seq assembly: BWA-MEM 2,350x coverage                               |
| EPI_ISL_427391 | İstanbul | Male   | 51  | 2020-04-13    | 2020-04-18     |                                             | Illumina Next Seq assembly: BWA-MEM 0.7.17.1 5,845x coverage                    |
| EPI_ISL_424366 | Kayseri  | Male   | 82  | 2020-03-17    | 2020-04-13     |                                             | Illumina Miseq assembly: Burrows-Wheeler aligner v07.17-r1188 26,200x coverage   |
| EPI_ISL_417413 | NA  | Female | 27 | 2020-03-17 | 2020-03-25 | NA | Nanopore MinION assembly: Geneious Prime 245X coverage |
|----------------|-----|--------|----|------------|------------|----|-----------------------------------------------------|
| EPI_ISL_435057 | Adiyaman | Male | 80 | 2020-04-09 | 2020-05-02 | NA | Oxford Nanopore MinION assembly: Geneious Prime 2020.1.2 40X coverage |
| EPI_ISL_437304 | Nevşehir | Female | 54 | 2020-03-26 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437305 | Kocaeli | Female | 66 | 2020-03-27 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437306 | Kocaeli | Male | 30 | 2020-03-27 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437307 | Mardin | Female | 19 | 2020-03-25 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437308 | Ankara | Male | 54 | 2020-03-25 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437309 | Ankara | Female | 58 | 2020-03-26 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437310 | Ankara | Male | NA | 2020-03-27 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437311 | Ankara | Male | 61 | 2020-03-27 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437312 | Kocaeli | Male | 52 | 2020-03-25 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437313 | Kocaeli | Male | 49 | 2020-03-27 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437314 | Ankara | Male | 62 | 2020-03-26 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437315 | Ankara | Female | 33 | 2020-03-26 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437316 | Denizli | Female | NA | 2020-03-25 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437317 | Ankara | Female | 58 | 2020-03-27 | 2020-05-08 | Patient travelled to Saudi Arabia | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437318 | Ankara | Male | 31 | 2020-03-19 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437319 | Kocaeli | Male | 47 | 2020-03-19 | 2020-05-08 | Patient travelled to Iran | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437320 | İstanbul | Female | 41  | 2020-03-19 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
|----------------|---------|--------|-----|------------|------------|----|-------------------------------------------------------------|
| EPI_ISL_437321 | İstanbul | Female | 25  | 2020-03-19 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437322 | Ankara  | Female | 62  | 2020-03-19 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437323 | İstanbul | Male   | 41  | 2020-03-19 | 2020-05-08 | Patient travelled to Taiwan | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437324 | İstanbul | Male   | 27  | 2020-03-19 | 2020-05-08 | Patient travelled to Iran | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437325 | İstanbul | Male   | 44  | 2020-03-19 | 2020-05-08 | Patient travelled to Iran | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437326 | İstanbul | Male   | 38  | 2020-03-19 | 2020-05-08 | Patient travelled to Iran | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437327 | Ağrı     | Female | 35  | 2020-03-19 | 2020-05-08 | Patient travelled to Iran | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437328 | Tekirdağ | Female | 35  | 2020-03-19 | 2020-05-08 | Patient travelled to Saudi Arabia | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437329 | Ankara   | Male   | 31  | 2020-03-19 | 2020-05-08 | Patient travelled to Saudi Arabia | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437330 | Tokat    | Male   | 20  | 2020-03-19 | 2020-05-08 | Health worker | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437331 | Ankara   | Female | 59  | 2020-03-25 | 2020-05-08 | Patient travelled to Saudi Arabia | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437332 | İstanbul | Male   | 50  | 2020-03-18 | 2020-05-08 | Patient travelled to Iran | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437333 | Ankara   | Male   | 64  | 2020-03-25 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437334 | Ankara   | Female | 64  | 2020-03-24 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
| EPI_ISL_437335 | Denizli  | Female | 79  | 2020-03-25 | 2020-05-08 | NA | Oxford Nanopore GridION assembly: Geneious Prime 1000X coverage |
**Supplementary Table 2.** Nucleotide substitutions present in 62 SARS-CoV-2 viral genomes from Turkey (submitted to GISAID between March 25th and May 22nd 2020) compared to the SARS-CoV-2 reference genome NC_045512.1. The nucleotide positions are given starting from 5’UTR. The star mark (*) indicates mutations that are only seen in the corresponding viral isolate.

| GISAID_sample accession ID | Nucleotide substitution at the given position | Corresponding viral gene | Corresponding viral gene product |
|----------------------------|-----------------------------------------------|---------------------------|---------------------------------|
| EPI_ISL_417413             |                                               |                           |                                 |
| *T580A                     | ORF1ab                                        | Leader protein            |                                 |
| *G779C                     |                                               |                           |                                 |
| *T946A                     |                                               |                           |                                 |
| *T1100G                    |                                               |                           |                                 |
| *C1101T                    |                                               |                           |                                 |
| *A1106T                    |                                               |                           |                                 |
| *A1119C                    |                                               |                           |                                 |
| *A1134T                    |                                               |                           |                                 |
| *G1156A                    |                                               |                           |                                 |
| *G1210A                    |                                               |                           |                                 |
| *C1225A                    |                                               |                           |                                 |
| *T1359C                    |                                               |                           |                                 |
| G1397A                     |                                               |                           |                                 |
| *C1420T                    |                                               |                           |                                 |
| *G1470A                    |                                               |                           |                                 |
| *G1472T                    |                                               |                           |                                 |
| *A1475C                    |                                               |                           |                                 |
| *G2250A                    |                                               |                           |                                 |
| C2455T                     |                                               |                           |                                 |
| *A2475T                    |                                               |                           |                                 |
| *G2549C                    |                                               |                           |                                 |
| *T2586A                    |                                               |                           |                                 |
| *G2591A                    |                                               |                           |                                 |
| *G2612C                    |                                               |                           |                                 |
| *G27151                    |                                               |                           |                                 |
| *A2922G                    |                                               |                           |                                 |
| C3117T                     |                                               |                           |                                 |
| *G3146C                    |                                               |                           |                                 |
| C3787T                     |                                               |                           |                                 |
| C4084T                     |                                               |                           |                                 |
| C7392T                     |                                               |                           |                                 |
| *C11232T                   |                                               |                           |                                 |
| *G11234A                   |                                               |                           |                                 |
| *C134761                   |                                               |                           |                                 |
| *C13492T                   |                                               |                           |                                 |
| *C142861                   |                                               |                           |                                 |
| *G14310A                   |                                               |                           |                                 |
| *T14394A                   |                                               |                           |                                 |
| *C14407A                   |                                               |                           |                                 |
| *G14430A                   |                                               |                           |                                 |
| *G14443T                   |                                               |                           |                                 |
| *T14682G                   |                                               |                           |                                 |
| *G14710A                   |                                               |                           |                                 |
| *T14740C                   |                                               |                           |                                 |
| *C14763A                   |                                               |                           |                                 |
| *G14773T                   |                                               |                           |                                 |
| *T14808A                   |                                               |                           |                                 |
| *G15101A                   |                                               |                           |                                 |
| *T15119A                   |                                               |                           |                                 |
| *G15958A                   |                                               |                           |                                 |
| *C19763A                   |                                               |                           |                                 |
| *T26396A                   |                                               |                           |                                 |
| *T26551C                   |                                               |                           |                                 |
| *C26753T                   |                                               |                           |                                 |
| C27103T                    |                                               |                           |                                 |
| G28109T                    |                                               |                           |                                 |
| T28688C                    |                                               |                           |                                 |
| G29742T                    |                                               |                           |                                 |
| Sequence ID | Position | ORF/Protein | Mutation | Protein Function  |
|------------|----------|-------------|----------|------------------|
| EPI_ISL_424366 | G1397A | ORF1ab | Nsp2 |  |
| EPI_ISL_424366 | G11083T | ORF1ab | Nsp6 |  |
| EPI_ISL_424366 | *G23876A | S | Spike glycoprotein |  |
| EPI_ISL_424366 | T26888C | N | Nucleocapsid phosphoprotein |  |
| EPI_ISL_424366 | *C29563T | | 3'UTR stem loop II like motif |  |
| EPI_ISL_427391 | C2113T | ORF1ab | Nsp2 |  |
| EPI_ISL_427391 | *C2997T | ORF1ab | Nsp3 |  |
| EPI_ISL_427391 | C23897I | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_427391 | C14408T | ORF1ab | Helicase |  |
| EPI_ISL_427391 | C18977T | ORF1ab | 3' to 5' exonuclease |  |
| EPI_ISL_427391 | A23403G | S | Spike glycoprotein |  |
| EPI_ISL_427391 | G25563T | ORF3a | ORF3a protein |  |
| EPI_ISL_428368 | C3037T | ORF1ab | Nsp3 |  |
| EPI_ISL_428368 | G1883T | ORF1ab | Nsp6 |  |
| EPI_ISL_428368 | *C23897I | ORF1ab | Nsp9 |  |
| EPI_ISL_428368 | C14408T | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_428368 | C17960T | ORF1ab | Helicase |  |
| EPI_ISL_428368 | C18977T | ORF1ab | 3' to 5' exonuclease |  |
| EPI_ISL_428368 | *G21452I | ORF1ab | 2'-O-Ribose methyltransferase |  |
| EPI_ISL_428368 | A23403G | S | Spike glycoprotein |  |
| EPI_ISL_428368 | G25563T | ORF3a | ORF3a protein |  |
| EPI_ISL_428346 | C3072T | ORF1ab | Nsp7 |  |
| EPI_ISL_428346 | C7962T | ORF1ab | Nsp8 |  |
| EPI_ISL_428346 | C14408T | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_428346 | C17960T | ORF1ab | Helicase |  |
| EPI_ISL_428346 | C18977T | ORF1ab | 3' to 5' exonuclease |  |
| EPI_ISL_428346 | *G21452I | ORF1ab | 2'-O-Ribose methyltransferase |  |
| EPI_ISL_428346 | A23403G | S | Spike glycoprotein |  |
| EPI_ISL_428346 | G25563T | ORF3a | ORF3a protein |  |
| EPI_ISL_428346 | C28899T | M | Membrane glycoprotein |  |
| EPI_ISL_428717 | C3072T | ORF1ab | Nsp7 |  |
| EPI_ISL_428717 | C7962T | ORF1ab | Nsp8 |  |
| EPI_ISL_428717 | C14408T | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_428717 | C17960T | ORF1ab | Helicase |  |
| EPI_ISL_428717 | C18977T | ORF1ab | 3' to 5' exonuclease |  |
| EPI_ISL_428717 | *G21452I | ORF1ab | 2'-O-Ribose methyltransferase |  |
| EPI_ISL_428717 | A23403G | S | Spike glycoprotein |  |
| EPI_ISL_428717 | G25563T | ORF3a | ORF3a protein |  |
| EPI_ISL_428719 | C3072T | ORF1ab | Nsp7 |  |
| EPI_ISL_428719 | C7962T | ORF1ab | Nsp8 |  |
| EPI_ISL_428719 | C14408T | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_428719 | C17960T | ORF1ab | Helicase |  |
| EPI_ISL_428719 | C18977T | ORF1ab | 3' to 5' exonuclease |  |
| EPI_ISL_428719 | *G21452I | ORF1ab | 2'-O-Ribose methyltransferase |  |
| EPI_ISL_428719 | A23403G | S | Spike glycoprotein |  |
| EPI_ISL_428719 | G25563T | ORF3a | ORF3a protein |  |
| EPI_ISL_428718 | C8782T | ORF1ab | Nsp4 |  |
| EPI_ISL_428718 | *G1813T | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_428718 | G28879A | N | Nucleocapsid phosphoprotein |  |
| EPI_ISL_428718 | G29742T | 3'UTR stem loop II like motif |  |
| EPI_ISL_428720 | C3072T | ORF1ab | Nsp7 |  |
| EPI_ISL_428720 | C7962T | ORF1ab | Nsp8 |  |
| EPI_ISL_428720 | C14408T | ORF1ab | RNA-dependent RNA polymerase |  |
| EPI_ISL_428720 | C18977T | ORF1ab | 3' to 5' exonuclease |  |
| EPI_ISL_428720 | A23403G | S | Spike glycoprotein |  |
| EPI_ISL_428720 | G25563T | ORF3a | ORF3a protein |  |
| EPI_ISL_428720 | C86735T | M | Membrane glycoprotein |  |
| EPI_ISL_428722 | C884T | ORF1ab | Nsp2 |
|---------------|-------|--------|------|
| G1397A        | ORF1ab| Nsp4   |      |
| G8625T        | ORF1ab| Nsp6   |      |
| G1054T        | ORF1ab| Nsp9   |      |
| GI3727       | ORF1ab| N      |      |
| G28688C       | N      | Nucleocapsid phosphoprotein |    |
| G927424       | N      | 3'UTR stem loop II like motif |    |

| EPI_ISL_428721 | C303T | ORF1ab | Nsp3 |
|---------------|-------|--------|------|
| C3237G        | ORF1ab| RNA-dependent RNA polymerase |    |
| C34405T       | ORF1ab| 3' to 5' exonuclease |    |
| C11083T       | ORF1ab| Nsp5   |      |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | ORF3a | ORF3a protein |    |
| G26735T       | M     | Membrane glycoprotein |    |

| EPI_ISL_428723 | C884T | ORF1ab | Nsp2 |
|---------------|-------|--------|------|
| C3237G        | ORF1ab| Nsp3   |      |
| C34405T       | ORF1ab| RNA-dependent RNA polymerase |    |
| C11083T       | ORF1ab| 3' to 5' exonuclease |    |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | ORF3a | ORF3a protein |    |
| G26735T       | M     | Membrane glycoprotein |    |

| EPI_ISL_428713 | G1397A | ORF1ab | Nsp2 |
|---------------|-------|--------|------|
| C3037T        | ORF1ab| Nsp3   |      |
| C14178T       | ORF1ab| RNA-dependent RNA polymerase |    |
| C18877T       | ORF1ab| 3' to 5' exonuclease |    |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | ORF3a | ORF3a protein |    |
| G26735T       | M     | Membrane glycoprotein |    |

| EPI_ISL_428712 | C303T | ORF1ab | Nsp3 |
|---------------|-------|--------|------|
| C3237G        | ORF1ab| RNA-dependent RNA polymerase |    |
| C34405T       | ORF1ab| 3' to 5' exonuclease |    |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | ORF3a | ORF3a protein |    |
| G26735T       | M     | Membrane glycoprotein |    |

| EPI_ISL_428715 | C303T | ORF1ab | Nsp3 |
|---------------|-------|--------|------|
| C3237G        | ORF1ab| RNA-dependent RNA polymerase |    |
| C34405T       | ORF1ab| 3' to 5' exonuclease |    |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | ORF3a | ORF3a protein |    |
| G26735T       | M     | Membrane glycoprotein |    |

| EPI_ISL_429870 | C884T | ORF1ab | Nsp2 |
|---------------|-------|--------|------|
| C3237G        | ORF1ab| Nsp3   |      |
| C34405T       | ORF1ab| RNA-dependent RNA polymerase |    |
| C11083T       | ORF1ab| 3' to 5' exonuclease |    |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | N     | Nucleocapsid phosphoprotein |    |

| EPI_ISL_429873 | C884T | ORF1ab | Nsp2 |
|---------------|-------|--------|------|
| C3237G        | ORF1ab| Nsp3   |      |
| C34405T       | ORF1ab| RNA-dependent RNA polymerase |    |
| C11083T       | ORF1ab| 3' to 5' exonuclease |    |
| A23403G       | S     | Spike glycoprotein |    |
| G25263T       | N     | Nucleocapsid phosphoprotein |    |
| EPI_ISL_429872 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| G1397A         | ORF1ab| Nsp4   |
| G8653T         | ORF1ab| Nsp6   |
| G11083T        | ORF1ab| Nsp6   |
| G29742T        | N     | Nucleocapsid phosphoprotein |

| EPI_ISL_429862 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| A30283G        | ORF1ab| endorRNAse |
| A23403G        | S     | Spike glycoprotein |

| EPI_ISL_429861 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| C18877T        | ORF1ab| 3’ to 5’ exonuclease |
| C22444T        | S     | Spike glycoprotein |
| G25263T        | ORF3a | ORF3a protein |
| C8853A         | N     | Nucleocapsid phosphoprotein |

| EPI_ISL_429864 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| A30283G        | S     | Spike glycoprotein |
| G25563T        | ORF3a | ORF3a protein |
| C26735T        | M     | Membrane glycoprotein |

| EPI_ISL_429865 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| C18877T        | ORF1ab| 3’ to 5’ exonuclease |
| A23403G        | S     | Spike glycoprotein |
| G25563T        | ORF3a | ORF3a protein |
| C26735T        | M     | Membrane glycoprotein |

| EPI_ISL_429868 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| C18877T        | ORF1ab| 3’ to 5’ exonuclease |
| A23403G        | S     | Spike glycoprotein |
| G25563T        | ORF3a | ORF3a protein |
| C26735T        | M     | Membrane glycoprotein |

| EPI_ISL_429867 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| C18877T        | ORF1ab| 3’ to 5’ exonuclease |
| A23403G        | S     | Spike glycoprotein |
| G25563T        | ORF3a | ORF3a protein |
| C26735T        | M     | Membrane glycoprotein |

| EPI_ISL_429869 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C30271         | ORF1ab| Nsp5   |
| C14408T        | ORF1ab| RNA-dependent RNA polymerase |
| C18877T        | ORF1ab| 3’ to 5’ exonuclease |
| A23403G        | S     | Spike glycoprotein |
| G25563T        | ORF3a | ORF3a protein |

| EPI_ISL_437304 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C22441         | S     | UTR    |
| C14408T        | ORF1ab| Nsp3   |
| C147021        | ORF1ab| 3C-like protease |
| G11083T        | ORF1ab| Nsp6   |
| G25263T        | S     | Spike glycoprotein |
| C26735T        | M     | Membrane glycoprotein |
| G25563T        | N     | Nucleocapsid phosphoprotein |

| EPI_ISL_437304 | C884T | ORF1ab | Nsp2 |
|----------------|-------|--------|------|
| C22441         | S     | UTR    |
| C14408T        | ORF1ab| Nsp3   |
| C147021        | ORF1ab| 3C-like protease |
| G11083T        | ORF1ab| Nsp6   |
| G25263T        | S     | Spike glycoprotein |
| C26735T        | M     | Membrane glycoprotein |
| G25563T        | N     | Nucleocapsid phosphoprotein |
| EPI_ISL_437305 |
|---|
| C241 | 5' UTR |
| C3037 | ORF1ab |
| C14408 | RNA-dependent RNA polymerase |
| A23403 | S Spike glycoprotein |
| G28881 | N Nucleocapsid phosphoprotein |
| G28882 | A |
| G28883 | C |

| EPI_ISL_437306 |
|---|
| G1397 | ORF1ab |
| G8653 | ORF1ab |
| G11083 | ORF1ab |
| A24030 | N Spike glycoprotein |
| L28688 | N Nucleocapsid phosphoprotein |
| G29742 | 3' UTR stem loop II like motif |

| EPI_ISL_437307 |
|---|
| *C1314 | ORF1ab |
| G1397 | ORF1ab |
| T5182 | ORF1ab |
| A23734 | ORF1ab |
| A23403 | RNA-dependent RNA polymerase |
| A25563 | ORF3a |

| EPI_ISL_437308 |
|---|
| C241 | 5' UTR |
| C3037 | ORF1ab |
| C14408 | RNA-dependent RNA polymerase |
| C18877 | ORF1ab |
| A23403 | S Spike glycoprotein |
| G25563 | ORF3a |

| EPI_ISL_437309 |
|---|
| C241 | 5' UTR |
| C3037 | ORF1ab |
| C14408 | RNA-dependent RNA polymerase |
| C22444 | S Spike glycoprotein |

| EPI_ISL_437310 |
|---|
| C241 | 5' UTR |
| C3037 | ORF1ab |
| C14408 | RNA-dependent RNA polymerase |
| C28854 | N Nucleocapsid phosphoprotein |

| EPI_ISL_437311 |
|---|
| C1397 | ORF1ab |
| T5182 | ORF1ab |
| G8653 | ORF1ab |
| G25563 | ORF3a |

| EPI_ISL_437312 |
|---|
| C241 | 5' UTR |
| C3037 | ORF1ab |
| C14408 | RNA-dependent RNA polymerase |
| A23874 | N Nucleocapsid phosphoprotein |

| EPI_ISL_437313 |
|---|
| G1397 | ORF1ab |
| T5182 | ORF1ab |
| G8653 | ORF1ab |
| G28854 | N Nucleocapsid phosphoprotein |

| EPI_ISL_437314 |
|---|
| G1397 | ORF1ab |
| T5182 | ORF1ab |
| G8653 | ORF1ab |
| G28854 | N Nucleocapsid phosphoprotein |
| Sample ID  | Mutation  | Position | Gene       | Protein       | Description                     |
|-----------|-----------|----------|------------|---------------|---------------------------------|
| EPI_ISL_437315 | C1059A   | ORF1ab   | Nsp2      | Spike glycoprotein |
| EPI_ISL_437316 | C241T   | 5' UTR   | Nsp3      | Spike glycoprotein |
| EPI_ISL_437317 | C241T   | 5' UTR   | Nsp4      | S Spike glycoprotein |
| EPI_ISL_437318 | C24T    | 5' UTR   | Nsp6      | S Spike glycoprotein |
| EPI_ISL_437319 | G1397A  | ORF1ab   | Nsp2      | 3' UTR stem loop II like motif |
| EPI_ISL_437320 | C228T   | 5' UTR   | Nsp4      | 3' UTR stem loop II like motif |
| EPI_ISL_437321 | C228T   | 5' UTR   | Nsp6      | 3' UTR stem loop II like motif |
| EPI_ISL_437322 | C241T   | 5' UTR   | Nsp6      | 3' UTR stem loop II like motif |
| EPI_ISL_437323 | C241T   | 5' UTR   | Nsp6      | 3' UTR stem loop II like motif |
| Reference | Variants | Function |
|-----------|----------|----------|
| EPI_ISL_437324 | C11083T | Spike glycoprotein |
| G1297A | ORF1ab | Nsp2 |
| G86351 | ORF1ab | Nsp4 |
| C10702T | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A79644 | N | Spike glycoprotein |
| C26549T | M | Membrane glycoprotein |
| T26551 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437325 | C11083T | Spike glycoprotein |
| G1297A | ORF1ab | Nsp2 |
| G86351 | ORF1ab | Nsp4 |
| C10702T | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A79644 | N | Spike glycoprotein |
| G11879 | M | Membrane glycoprotein |
| T26551 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437326 | C2411T | 5' UTR |
| C11083T | ORF1ab | Nsp2 |
| C10702T | ORF1ab | Nsp4 |
| A79644 | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A14408T | ORF1ab | RNA-dependent RNA polymerase |
| C22443 | N | Spike glycoprotein |
| G25563 | M | Membrane glycoprotein |
| A26720 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437327 | C2411T | 5' UTR |
| C11083T | ORF1ab | Nsp2 |
| C10702T | ORF1ab | Nsp4 |
| A79644 | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A14408T | ORF1ab | RNA-dependent RNA polymerase |
| C22443 | N | Spike glycoprotein |
| G25563 | M | Membrane glycoprotein |
| A26720 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437328 | C2411T | 5' UTR |
| C11083T | ORF1ab | Nsp2 |
| C10702T | ORF1ab | Nsp4 |
| A79644 | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A14408T | ORF1ab | RNA-dependent RNA polymerase |
| C22443 | N | Spike glycoprotein |
| G25563 | M | Membrane glycoprotein |
| A26720 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437329 | C2411T | 5' UTR |
| C11083T | ORF1ab | Nsp2 |
| C10702T | ORF1ab | Nsp4 |
| A79644 | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A14408T | ORF1ab | RNA-dependent RNA polymerase |
| C22443 | N | Spike glycoprotein |
| G25563 | M | Membrane glycoprotein |
| A26720 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437330 | C2411T | 5' UTR |
| C11083T | ORF1ab | Nsp2 |
| C10702T | ORF1ab | Nsp4 |
| A79644 | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A14408T | ORF1ab | RNA-dependent RNA polymerase |
| C22443 | N | Spike glycoprotein |
| G25563 | M | Membrane glycoprotein |
| A26720 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437331 | C2411T | 5' UTR |
| C11083T | ORF1ab | Nsp2 |
| C10702T | ORF1ab | Nsp4 |
| A79644 | ORF1ab | 3C-like proteinase |
| G10856 | ORF1ab | Nsp6 |
| A14408T | ORF1ab | RNA-dependent RNA polymerase |
| C22443 | N | Spike glycoprotein |
| G25563 | M | Membrane glycoprotein |
| A26720 | N | Nucleocapsid phosphoprotein |
| L28686 | N | 3' UTR stem loop II like motif |
| G297421 | N | 5' UTR |
| EPI_ISL_437333 | ORF1ab | Nsp3  |
|----------------|--------|-------|
| C3037T        | ORF1ab | RNA-dependent RNA polymerase |
| C144081       | ORF1ab | EndoRNase |
| *T13102C      | ORF1ab | Spike glycoprotein |
| G28637G       | S      | Membrane glycoprotein |
| G26549T       | M      | Nucleocapsid phosphoprotein |
| G28881A       | N      | |
| G28883C       |       | |

| EPI_ISL_437334 | ORF1ab | Nsp3  |
|----------------|--------|-------|
| C228T         | 5' UTR |       |
| G1397A        | ORF1ab | Nsp2  |
| C3037T        | ORF1ab | Nsp3  |
| C3903T        | ORF1ab | Nsp4  |
| C14408T       | ORF1ab | Nsp6  |
| C19484T       | ORF1ab | RNA-dependent RNA polymerase |
| G26720C       | M      | Membrane glycoprotein |
| T28688C       | N      | Nucleocapsid phosphoprotein |
| T28835C       |       | 3'UTR stem loop II like motif |
| G29742T       |       | |

| EPI_ISL_437335 | ORF1ab | Nsp3  |
|----------------|--------|-------|
| C2328T        | 5' UTR |       |
| C3037T        | ORF1ab | Nsp3  |
| C3903T        | ORF1ab | Nsp3  |
| C14408T       | ORF1ab | RNA-dependent RNA polymerase |
| A25304G       | S      | Spike glycoprotein |
| C26549T       | M      | Membrane glycoprotein |
| *A27354G      | ORF6   | ORF6 protein |