

1 **Abstract**

2 As the underlying pathogen for the COVID-19 pandemic that has affected tens of
3 millions of lives worldwide, SARS-CoV-2 and its mutations are among the most urgent
4 research topics worldwide. Mutations in the virus genome can complicate attempts at
5 accurate testing or developing a working treatment for the disease. Furthermore,
6 because the virus uses its own proteins to replicate its genome, rather than host proteins,
7 mutations in the replication proteins can have cascading effects on the mutation load of
8 the virus genome. Due to the global, rapidly developing nature of the COVID-19
9 pandemic, local demographics of the virus can be difficult to accurately analyze and
10 track, disproportionate to the importance of such information. Here, we analyzed
11 available, high-quality genome data of SARS-CoV-2 isolates from Turkey and
12 identified their mutations, in comparison to the reference genome, to understand how
13 the local mutasome compares to the global genomes. Our results indicate that viral
14 genomes in Turkey has one of the highest mutation loads and certain mutations are
15 remarkably frequent compared to global genomes. We also made the data on Turkey
16 isolates available on an online database to facilitate further research on SARS-CoV-2
17 mutations in Turkey.

18 **Keywords: COVID-19, SARS-CoV-2, coronavirus, genome analysis, mutation**
19 **profiling, Turkey, database**

20

21 **1. Introduction**

22 Coronavirus disease 2019 (COVID-19) is an ongoing pandemic, characterized by long-
23 term respiratory damage and slow onset fever, and caused by the SARS-CoV-2
24 betacoronavirus. The virus was first observed in human patients in late 2019, in the

1 Wuhan province of China, and soon after showed the capacity for human-to-human
2 transmission (Chan et al. 2020, Riou and Althaus 2020). As of 17 August 2020 there are
3 over 21 million confirmed cases and 767,158 recorded deaths. In addition to the
4 immediately apparent symptoms, its long term effect on humans are still topics of
5 research (Kochi et al. 2020, Lee et al. 2020, Li et al. 2020, Zhu et al. 2020). Because
6 COVID-19 is a highly transmissible disease with a capacity for asymptomatic
7 transmission (Wong et al. 2020), understanding the disease, as well as its underlying
8 pathogen and its routes of transmissions is a high priority topic. As a result, extant
9 databases on viral pathogens, such as GISAID (Elbe and Buckland-Merrett 2017) and
10 NextStrain (Hadfield et al. 2018) have become vital resources for researchers who seek
11 to track the evolution of the virus during its transmissions.

12 SARS-CoV-2 has a single-stranded RNA genome that codes for the proteins
13 responsible for its own replication, many of which are produced via cleavage of the
14 Orf1ab polyprotein, the largest gene on the genome. Therefore, mutations in the SARS-
15 CoV-2 genome can lead to cascading effects by reducing the fidelity of subsequent
16 replication cycles. Key proteins in the RNA replication complex include nsp7, 8, and
17 12 (also known as RNA dependent RNA polymerase or RdRp), which together form the
18 core polymerase complex (Kirchdoerfer and Ward 2019, Peng et al. 2020), as well as
19 nsp14, a dual function protein which joins the larger replication complex as a 3'-5'
20 error-correcting exonuclease (Subissi et al. 2014, Romano et al. 2020). Our previous
21 findings show that frequently observed mutations in both nsp12 and nsp14 are
22 associated with an increase in mutation density in the SARS-CoV-2 genome (Eskier et
23 al. 2020c, 2020a, 2020b).

1 In this study, we aimed to analyze the current mutasome of SARS-CoV-2 in Turkey,
2 with three main questions in mind: (i) are there any key reoccurring mutations observed
3 in a large number of isolates?; (ii) how does the distribution of mutations among
4 isolates compare to other regions in the world; and finally, (iii) are there any mutations
5 observed in Turkey but not the rest of the world? We focused on the latter two questions
6 in particular, with an emphasis on mutations of interest previously described in the
7 literature. Our findings reveal the presence of three main clades of SARS-CoV-2 in
8 Turkey, roughly analogous to 19A, 20A, and 20B as described in NextStrain, with a
9 preponderance of high mutability variants (Eskier et al. 2020c, 2020a, 2020b) compared
10 to international isolates. Furthermore, we identified several frequently recurrent,
11 previously uncharacterized variants in Turkey isolates not observed in isolates from
12 other countries, which can serve as potential candidates for validation and study.
13 Furthermore, we collected our analysis of Turkey isolates in a regularly maintained and
14 updated database, which we hope will serve as a potential resource for future research
15 on the local mutasome of SARS-CoV-2.

16 **2. Materials and Methods**

17 **2.1. Genome sequence filtering, retrieval, and preprocessing**

18 SARS-CoV-2 isolate genome sequences and the corresponding metadata were obtained
19 from the GISAID EpiCoV database on 28 July 2020⁴. These sequences were filtered for
20 location to limit our database to isolates with the location “Europe / Turkey”, which
21 resulted in 180 isolate sequences. We applied further quality filters, including selecting
22 only isolates obtained from human hosts (excluding environmental samples and animal
23 hosts), those sequenced for the full length of the genome (sequence size of 29 kb or
24 greater), and those with high coverage for the reference genome ($< 1\%$ N content, $<$

1 0.05% unique mutations, no unverified indel mutations), which further narrowed down
2 the list to 166 isolates. To ensure alignment accuracy, as characters that are not one of
3 A, C, G, T, or N would not be aligned according to potential biological meanings of the
4 alternative characters, all nonstandard unverified nucleotide masking was changed to N,
5 using the Linux *sed* command, and the isolates were aligned against the SARS-CoV-2
6 reference genome using the MAFFT (v7.450) alignment software (Katoh et al. 2002).
7 Variant sites in the isolates were annotated using snp-sites (2.5.1), bcftools (1.10.2)⁵,
8 and ANNOVAR (release date 24 October 2019) software (Wang et al. 2010, Page et al.
9 2016), to identify whether a given mutation was synonymous or nonsynonymous. In
10 addition, the 5' untranslated region of the genome (bases 1-265) and the 100 nucleotides
11 at the 3' end were removed from the alignment and annotation files due to a high
12 number of gaps and unidentified nucleotides.

13 **2.2. Development of the database and user interfaces**

14 The genome data is stored using the MariaDB 10.3.22 database installed on Debian
15 Linux 10 operating system. For web application, the genome data is visualized on the
16 map using jVectorMap with HTML 5 and Ajax web development techniques, using the
17 Django 3.0.5. framework and Python 3.7.3 programming language. A modified version
18 of TreeTime, an open-source phylogenetic analysis software, is used to create the
19 phylogenetic tree (Sagulenko et al. 2018).

20 **3. Results**

21 **3.1. The distributions of mutations across isolates in Turkey**

22 Our analysis of the genome sequences of 166 isolates from Turkey revealed 258 distinct
23 mutations across the isolates, 87 of which are observed in multiple isolates, and 43 of
24 them are found in at least five isolates (hereafter referred to as recurring mutations). 19

1 of the 43 recurring mutations are nonsynonymous, 21 are synonymous, and 3 are found
2 outside of coding regions. C>T transitions are the most common, comprising over half
3 of the mutations, consistent with previous international findings on C>U
4 hypermutations in SARS-CoV-2 (Simmonds 2020). The most commonly seen
5 mutations are 3037 C>T, 14408 C>T, and 23403 A>G, observed together in 139 of the
6 isolates, with one singleton instance of 23403 A>G, also consistent with previous
7 findings (Pachetti et al. 2020, Yin 2020). Orf1ab mutations are the most common,
8 comprising 23 of the recurring mutations, consistent with the size of the gene, as Orf1ab
9 makes up two thirds of the SARS-CoV-2 genome. Orf9 (nucleocapsid; N) gene has the
10 second highest number of recurring mutations ($n = 7$, however, 3 of them are block
11 mutations of 28881-28883 trinucleotide), followed by Orf5 (membrane; M) and S genes
12 ($n = 5$) (Table 1).

13 To identify which of the recurring mutations are stronger indicators of Turkey
14 genotype, we compared their frequency in the isolate population from Turkey to
15 frequencies in other geographical regions, using a metric of mutation instance per
16 sequenced isolate. To eliminate the potential confounding effect of earlier isolates
17 having a lower number of mutations on average, and different regions having started
18 sequencing efforts in different timetables, we selected isolates sequenced after the day
19 when each region of interest had at least ten isolates sequenced. As Turkey was the
20 latest region to have the required number of isolates (19 March 2020), it was used as the
21 filtering metric. Four of the recurring mutations were found only in Turkey isolates, and
22 six more were not recurring mutations in other regions. Therefore, we focused the
23 comparisons on the remaining 33 recurring mutations. We identified the percentages of
24 these mutations in each region (Africa, Asia, Europe, North America, Oceania, South

1 America), as well as worldwide totals, excluding isolates from Turkey, where
2 applicable, and compared them to the corresponding percentages in Turkey (Table 2).
3 With two exceptions (241 C>T and 20268 A>G), all of the mutations have higher
4 percentages in Turkey compared to worldwide percentages. In addition, isolates from
5 Turkey comprise over 50% of the isolates worldwide carrying six of the mutations (228
6 C>T, 8326 C>T, 12809 C>T, 13620 C>T, 14724 C>T, and 27703 C>T), and 20 of the
7 mutations have higher percentages in Turkey than any other region. Among these six
8 mutations, 228 C>T was first detected in Canada on March 11 and in Turkey (Istanbul)
9 on March 18⁶. 8326 C>T was first isolated from a patient in Taipei on March 19 and
10 only two other cases was reported (UK and Denmark) before the first case in Turkey
11 (Kars) on April 29. Thirteen more cases with 8326 C>T in the same city implicates
12 local transmission, and with 8 cases reported on the most recent update (July 15), this
13 particular mutation is a candidate for even further spread, assuming that the isolates
14 sequenced were randomly selected across the infected population, instead of all being
15 selected from a known cluster of patients. 12809 C>T mutation was first reported in an
16 isolate from Washington/USA collected on March 14, which spread to five more states
17 by the end of the month; however, only 4 more cases were reported afterwards, the last
18 being one on April 27. The first isolate from Turkey with 12809 C>T mutation was
19 collected on April 13 in Istanbul (EPI_ISL_480230). Being present in only three other
20 countries with one isolate each (UK, India, Australia), in addition to the USA, this
21 mutation was very likely an introduction from the USA. Since then, 20 more cases in
22 Istanbul and 3 more cases in Kars were reported, all in April and May. 13620 C>T
23 mutation was first reported in an isolate in Italy on March 5 and later in South Africa,
24 USA, Denmark, Belgium, Luxembourg and Singapore by April 6. However, only three

1 more cases were reported for the rest of April, two in Italy on April 12, and one in USA
2 on April 28. The first case with the same mutation in Turkey was reported in Kars on
3 May 17 and has been reported in 18 more cases, all in the same city. Initial phylogenetic
4 analysis does not support introduction of this mutation from abroad, however, limited
5 sampling makes it difficult to reach a definitive conclusion. 13620 C>T, 14724 C>T,
6 and 27703 G>T mutations are linked in SARS-CoV-2 genomes from Turkey, all from a
7 single city (Kars), suggesting a founder effect and local transmission. It bears noting
8 that each of the four mutations exclusively found as recurring in Turkey are limited to a
9 single batch of isolates obtained by a single center, therefore pending verification.

10 Afterwards, we sought to understand how the mutation load of the isolates in Turkey
11 compare to distributions in other regions. Using our previous date filter, we calculated
12 the number of single nucleotide variants (SNVs) per isolate in each region (Table 3).
13 Turkey had the highest number of SNVs per isolate, followed by South America. In
14 comparison, Africa, another region which started sequencing efforts later than the other
15 regions, had a mean SNV number lower than that of Asia, the region with the earliest
16 sequences available, implying that the mutation numbers are strongly influenced by
17 other factors in addition to the date of introduction of the virus to the region. We also
18 compared the number of SNVs per isolate in each region per gene, normalized by
19 kilobase of gene region (Table 4). Turkey had the most SNVs of any region in Orf1ab,
20 M, and Orf7a genes, with Orf7a having more than three times as many SNVs as any
21 other region.

22 **3.2. Database implementation**

23 Data regarding Turkey isolates are available as a database comprising an interactive
24 phylogenetic tree of the isolates, a geographical heatmap of sequenced isolates, and

1 tables for both the mutasome of individual isolates, and summaries of the mutations
2 observed in the isolates. The phylogenetic tree can be viewed both in real time and
3 divergence time, and colored according to nucleotide of interest, location, or sequencing
4 date. The tables are generated using the sequencing metadata available from GISAID as
5 well as ANNOVAR variant annotation tables. We aim to regularly validate and update
6 the database as new sequences are made available. The database is freely accessible at
7 <http://covid19.ibg.edu.tr>. Future plans include implementation of NextStrain clade and
8 branch information in the phylogenetic tree to aid the user in comparisons with
9 international sequencing data.

10 **4. Discussion**

11 COVID-19 has been causing tremendous challenges for clinicians, healthcare systems,
12 societies, and governments, and has required development of novel approaches to fight
13 the pandemic. With an unpredictable future course for the ongoing pandemic, close
14 monitoring and characterization of mutations has emerged as top priorities for better
15 understanding of possible genotype-phenotype relations, and therefore better
16 management of healthcare efforts.

17 Mutations in any viral infection, especially those that have crossed interspecies barriers,
18 have to be considered in the context of natural selection. As the evolution of a virus will
19 likely affect its fitness in a new host, any attempts against such an infection have to
20 consider the causal relationships between genomic variances and the spread of the virus.

21 Previous studies suggest that the selective pressure on mutations in SARS-CoV-2 in
22 human hosts are largely confined to modest positive selection, with very little purifying
23 selection, due to the short span of the pandemic, and that most of the positive selection
24 have occurred in previous hosts (MacLean et al. 2020). Therefore, any investigation of

1 the mutations will need to consider most of the mutations have to be beneficial or
2 neutral to create true strains of the virus. A comprehensive analysis by Jungreis et al.
3 showed that SARS-Cov-2 mutations are excluded from the evolutionarily conserved
4 amino acid residues and nucleotides, and the authors concluded both synonymous and
5 non-synonymous mutations are under purifying selection (Jungreis et al. 2020).
6 Therefore, not only the non-synonymous mutations, but also the synonymous ones
7 should be considered as potentially functional.

8 Many studies already provided lines of evidence that supports a role for the S D614G
9 mutation in increased infectivity and likely in transmissibility of SARS-CoV-2 (Korber
10 et al. 2020, Daniloski et al. 2020). It is possible that new mutations that affect viral
11 behavior may arise, and therefore emergence and spreading of such mutations should be
12 monitored closely. However, with tens of millions affected worldwide, monitoring of
13 every single mutation is a challenging task. We believe that our database will provide a
14 valuable and practical resource for researchers in Turkey, as well as in other countries,
15 to track the spread of SARS-CoV-2 mutations in Turkey.

16 Our findings show the viral isolates in Turkey have accumulated a higher number of
17 mutations compared to other regions on average, even after normalizing for the isolates
18 sequenced earlier during the pandemic having accumulated fewer mutations.

19 Furthermore, it has more mutations in the Orf1ab gene, which produces the polyprotein
20 that is cleaved into the mature peptides responsible for viral replication, than any other
21 region. In addition, it has the third highest number of mutations in the S gene, which is
22 responsible for the viral infection of the cells. As these two genes have the highest
23 potential impact on the replication and transmission cycle of the virus, a higher
24 mutation density in these genes can lead to an accelerated mutation rate. Of note, the

1 18877 C>T mutation in nsp14, the 3'-5' exonuclease responsible for error correction
2 during genomic replication, has the second highest frequency in Turkey of any country⁶.
3 Our previous study (Eskier et al. 2020b) shows a strong correlation between increased
4 mutation density and the 18877 C>T mutation, which might be a potential reason for
5 Turkey's increased SNV average per isolate.

6 Two groups of mutations we identified that is worth further attention are the 3037 C>T,
7 14408 C>T, 23403 A>G haplotype, and the 28881-28883 block mutation. Both of these
8 groups of mutations are found almost exclusively together, both in Turkey, and
9 worldwide. In both cases, Turkey has a higher incidence of mutations in these groups
10 than worldwide averages, and four of the major regions (Asia, Europe, North America,
11 Oceania). We previously found that the 14408 C>T and 23403 A>G mutations, when
12 occurring together, are strongly associated with increased mutation density over time
13 (Eskier et al. 2020a), and the prevalence of both these mutations and the 18877 C>T
14 mutation in Turkey isolates may further contribute to a variant-rich mutation landscape
15 (Eskier et al. 2020b). 28881-28883 GGG>AAC is found on the N gene, whose product
16 is responsible for packaging the genome into newly produced virions in cells, and
17 regulating host cell response (McBride et al. 2014). The mutation disrupts an SR-rich
18 motif in the nucleocapsid protein, which was found to cause reduced transmissibility in
19 SARS-CoV, a similar betacoronavirus with high homology to SARS-CoV-2 (Tylor et
20 al. 2009, Ayub 2020). It is not clear whether the mutation groups are selected together
21 and show homoplastic recurrence across isolates, or if they are a result of strong founder
22 effect.

23 A major concern when analyzing the isolate sequences from Turkey is the limited
24 nature of the data. The sequences are few in number, and their geographical and

1 temporal distributions are highly skewed, leading to difficulty in understanding the
2 transmission routes of the virus across the country. Furthermore, new sequences are
3 often made available in large batches by the centers, which further introduces bias to the
4 samples by potentially generating sequencing or assembly artifacts to the sequences.
5 Unless verified by multiple centers, in multiple batches, or by other experimental
6 methods, caution is required when studying these mutations. As more genomes are
7 sequenced, a more clear picture of the SARS-CoV-2 mutasome in Turkey will emerge
8 and we will likely be able to draw more solid conclusions.

9 Finally, it should be noted that mutational profiles of viral genomes may determine
10 whether infected patients will develop lasting immunity and remain protected from re-
11 infection. Although exposure to SARS-CoV-2 protected rhesus macaques from re-
12 infection with the same strain of virus (Deng et al. 2020), there are questions still
13 remaining to be answered related to whether each recovered patient will have lasting
14 immunity. Recent news within days reported that four patients from Hong Kong,
15 Belgium, the Netherlands, and USA, who had earlier recovered from COVID-19 has
16 been re-infected, with a different strain of SARS-CoV-2 than the original infection⁷
17 (Tillett et al. 2020). In support of this observation, an earlier study reported that
18 convalescent plasma from some of the COVID-19 patients showed reduced neutralizing
19 activity against pseudoviruses with D614G mutation in culture environment (Hue et al.
20 2020). We do not have a clear understanding of the viral determinants of lasting
21 immunity to SARS-CoV-2, however, it seems that certain viral proteins may be more
22 critical than others, based on analyses of patient plasma samples. Grifoni et al.
23 suggested that M, Spike and N proteins are the major determinants of CD4+ response,
24 with additional responses to nsp3, nsp4, ORF3a and ORF8 (Grifoni et al. 2020).

1 Hachim et al. showed that ORF8, ORF3b and N proteins of SARS-CoV-2 elicited the
2 strongest specific antibody responses in infected patients (Hachim et al. 2020). It is
3 plausible that certain mutations within these proteins affect the immune response,
4 however, it remains to be explored whether any of the mutations common or more
5 frequently seen in Turkish isolates have any effect on the immune response.

6 ¹ World Health Organization (1948). WHO timeline – COVID-19 [online]. Website:
7 <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19> [accessed:
8 17 August 2020].

9 ² The GISAID Initiative (2008). GISAID [online]. Website: <https://www.gisaid.org/>
10 [accessed: 27 August 2020].

11 ³ Nextstrain (2015). Nextstrain [online]. Website: <https://www.nextstrain.org> [accessed:
12 27 August 2020].

13 ⁴ The GISAID Initiative (2008). EpiCoV database [online]. Website:
14 <https://www.epicov.org/epi3/frontend> [accessed 28 July 2020].

15 ⁵ GitHub (2007). Bcftools repository [online]. Website:
16 <https://github.com/samtools/bcftools> [accessed 20 August 2020].

17 ⁶ China National Center for Bioinformation (2009). 2019 Novel Coronavirus Resource
18 (2019nCoV) [online]. Website: <https://bigd.big.ac.cn/ncov/variation/annotation/>
19 [accessed 28 August 2020].

1 ⁷ Euronews (1993). Euronews [online]. Website:
2 [https://www.euronews.com/2020/08/25/two-cases-of-covid-19-reinfection-reported-in-](https://www.euronews.com/2020/08/25/two-cases-of-covid-19-reinfection-reported-in-europe)
3 [europe](https://www.euronews.com/2020/08/25/two-cases-of-covid-19-reinfection-reported-in-europe) [accessed 28 August 2020].

4 **References**

5 Ayub M (2020) Reporting Two SARS-CoV-2 Strains Based on A Unique
6 Trinucleotide-Bloc Mutation and Their Potential Pathogenic Difference.
7 <https://doi.org/10.20944/preprints202004.0337.v1>

8 Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, et al. (2020) A familial cluster of
9 pneumonia associated with the 2019 novel coronavirus indicating person-to-
10 person transmission: a study of a family cluster. *Lancet* (London, England) 395:
11 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)

12 Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G, tenOever BR, Sanjana NE.
13 (2020) The Spike D614G mutation increases SARS-CoV-2 infection of multiple
14 human cell types. *bioRxiv*: 0.1101/2020.06.14.151357 doi:
15 <https://doi.org/10.1101/2020.06.14.151357>

16
17 Deng W, Bao L, Liu J, et al. Primary exposure to SARS-CoV-2 protects against
18 reinfection in rhesus macaques. (2020) *Science* 369(6505):818-823.
19 doi:10.1126/science.abc5343

20 Elbe S, Buckland-Merrett G (2017) Data, disease and diplomacy: GISAID’s innovative
21 contribution to global health. *Global Challenges* 1: 33–46.
22 <https://doi.org/10.1002/gch2.1018>

- 1 Eskier D, Suner A, Karakülah G, Oktay Y (2020a) Mutation density changes in SARS-
2 CoV-2 are related to the pandemic stage but to a lesser extent in the dominant
3 strain with mutations in spike and RdRp. Gillespie J (Ed). PeerJ 8: e9703.
4 <https://doi.org/10.7717/peerj.9703>
- 5 Eskier D, Suner A, Oktay Y, Karakülah G (2020b) Mutations of SARS-CoV-2 nsp14
6 exhibit strong association with increased genome-wide mutation load. bioRxiv:
7 2020.08.12.248732. <https://doi.org/10.1101/2020.08.12.248732>
- 8 Eskier D, Karakülah G, Suner A, Oktay Y (2020c) RdRp mutations are associated with
9 SARS-CoV-2 genome evolution. PeerJ 8: e9587.
10 <https://doi.org/10.7717/peerj.9587>
- 11 Grifoni A, Weiskopf D, Ramirez SI, et al. (2020) Targets of T Cell Responses to SARS-
12 CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed
13 Individuals. Cell 181(7):1489-1501.e15. doi:10.1016/j.cell.2020.05.015
14
- 15 Hachim A, Kavian N, Cohen CA, et al. (2020) ORF8 and ORF3b antibodies are
16 accurate serological markers of early and late SARS-CoV-2 infection [published
17 online ahead of print, 2020 Aug 17] [published correction appears in Nat
18 Immunol. 2020 Aug 27;:]. Nat Immunol. 10.1038/s41590-020-0773-7.
19 doi:10.1038/s41590-020-0773-7
20
- 21 Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P,
22 Bedford T, Neher RA (2018) Nextstrain: real-time tracking of pathogen

1 evolution. *Bioinformatics* 34: 4121–4123.
2 <https://doi.org/10.1093/bioinformatics/bty407>

3 Hu J, He CL, Gao QZ, Zhang GJ, et al. (2020) D614G mutation of SARS-CoV-2 spike
4 protein enhances viral infectivity. *bioRxiv*: 10.1101/2020.06.20.161323
5 <https://doi.org/10.1101/2020.06.20.161323>
6

7 Jungreis I, Sealfon R, Kellis M. (2020) Sarbecovirus comparative genomics elucidates
8 gene content of SARS-CoV-2 and functional impact of COVID-19 pandemic
9 mutations. Preprint. *bioRxiv*:2020.06.02.130955. Published 2020 Jun 3.
10 [doi:10.1101/2020.06.02.130955](https://doi.org/10.1101/2020.06.02.130955)
11

12 Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid
13 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids*
14 *Research* 30: 3059–3066. <https://doi.org/10.1093/nar/gkf436>

15 Kirchdoerfer RN, Ward AB (2019) Structure of the SARS-CoV nsp12 polymerase
16 bound to nsp7 and nsp8 co-factors. *Nature Communications* 10: 2342.
17 <https://doi.org/10.1038/s41467-019-10280-3>

18 Kochi AN, Tagliari AP, Forleo GB, Fassini GM, Tondo C (2020) Cardiac and
19 arrhythmic complications in patients with COVID-19. *Journal of Cardiovascular*
20 *Electrophysiology* 31: 1003–1008. <https://doi.org/10.1111/jce.14479>

21 Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, et al., Group on behalf of the
22 SC-19 G, LaBranche CC, Montefiori DC. (2020) Spike mutation pipeline

1 reveals the emergence of a more transmissible form of SARS-CoV-2. bioRxiv:
2 10.1101/2020.04.29.069054 <https://doi.org/10.1101/2020.04.29.069054>
3

4 Lee I-C, Huo T-I, Huang Y-H (2020) Gastrointestinal and liver manifestations in
5 patients with COVID-19. *Journal of the Chinese Medical Association: JCMA*
6 83: 521–523. <https://doi.org/10.1097/JCMA.0000000000000319>

7 Li Y-C, Bai W-Z, Hashikawa T (2020) The neuroinvasive potential of SARS-CoV2
8 may play a role in the respiratory failure of COVID-19 patients. *Journal of*
9 *Medical Virology* 92: 552–555. <https://doi.org/10.1002/jmv.25728>

10 MacLean OA, Lytras S, Weaver S, Singer JB, Boni MF, Lemey P, Pond SLK,
11 Robertson DL (2020) Natural selection in the evolution of SARS-CoV-2 in bats,
12 not humans, created a highly capable human pathogen. bioRxiv:
13 2020.05.28.122366. <https://doi.org/10.1101/2020.05.28.122366>

14 McBride R, Van Zyl M, Fielding BC (2014) The Coronavirus Nucleocapsid Is a
15 Multifunctional Protein. *Viruses* 6: 2991–3018.
16 <https://doi.org/10.3390/v6082991>

17 Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C,
18 Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R (2020) Emerging
19 SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA
20 polymerase variant. *Journal of Translational Medicine* 18: 179.
21 <https://doi.org/10.1186/s12967-020-02344-6>

- 1 Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR (2016)
2 SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments.
3 Microbial Genomics, 2. Available from:
4 <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.00>
5 0056.
- 6 Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, Wang Q, Sun Y, Fan Z, Qi J, Gao
7 GF, Shi Y (2020) Structural and Biochemical Characterization of the nsp12-
8 nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. Cell Reports 31:
9 107774. <https://doi.org/10.1016/j.celrep.2020.107774>
- 10 Riou J, Althaus CL (2020) Pattern of early human-to-human transmission of Wuhan
11 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro
12 Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European
13 Communicable Disease Bulletin 25. <https://doi.org/10.2807/1560->
14 7917.ES.2020.25.4.2000058
- 15 Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R (2020) A Structural View of
16 SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and
17 Final Capping. Cells 9. <https://doi.org/10.3390/cells9051267>
- 18 Sagulenko P, Puller V, Neher RA (2018) TreeTime: Maximum-likelihood
19 phylodynamic analysis. Virus Evolution 4. <https://doi.org/10.1093/ve/vex042>
- 20 Simmonds P (2020) Rampant C->U hypermutation in the genomes of SARS-CoV-2
21 and other coronaviruses – causes and consequences for their short and long

1 evolutionary trajectories. bioRxiv: 2020.05.01.072330.
2 <https://doi.org/10.1101/2020.05.01.072330>

3 Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E,
4 Snijder EJ, Canard B, Imbert I (2014) One severe acute respiratory syndrome
5 coronavirus protein complex integrates processive RNA polymerase and
6 exonuclease activities. *Proceedings of the National Academy of Sciences* 111:
7 E3900–E3909. <https://doi.org/10.1073/pnas.1323705111>

8 Tillett R, Sevinsky J, Hartley P, Kerwin H, Crawford N, Gorzalski A, Laverdure C,
9 Verma S, Rossetto C, Farrell M, Jackson D, Pandori M, van Hooser S. Genomic
10 Evidence for a Case of Reinfection with SARS-CoV-2. *THELANCETID-D-20-*
11 *05376*, Available at SSRN: <https://ssrn.com/abstract=3681489>

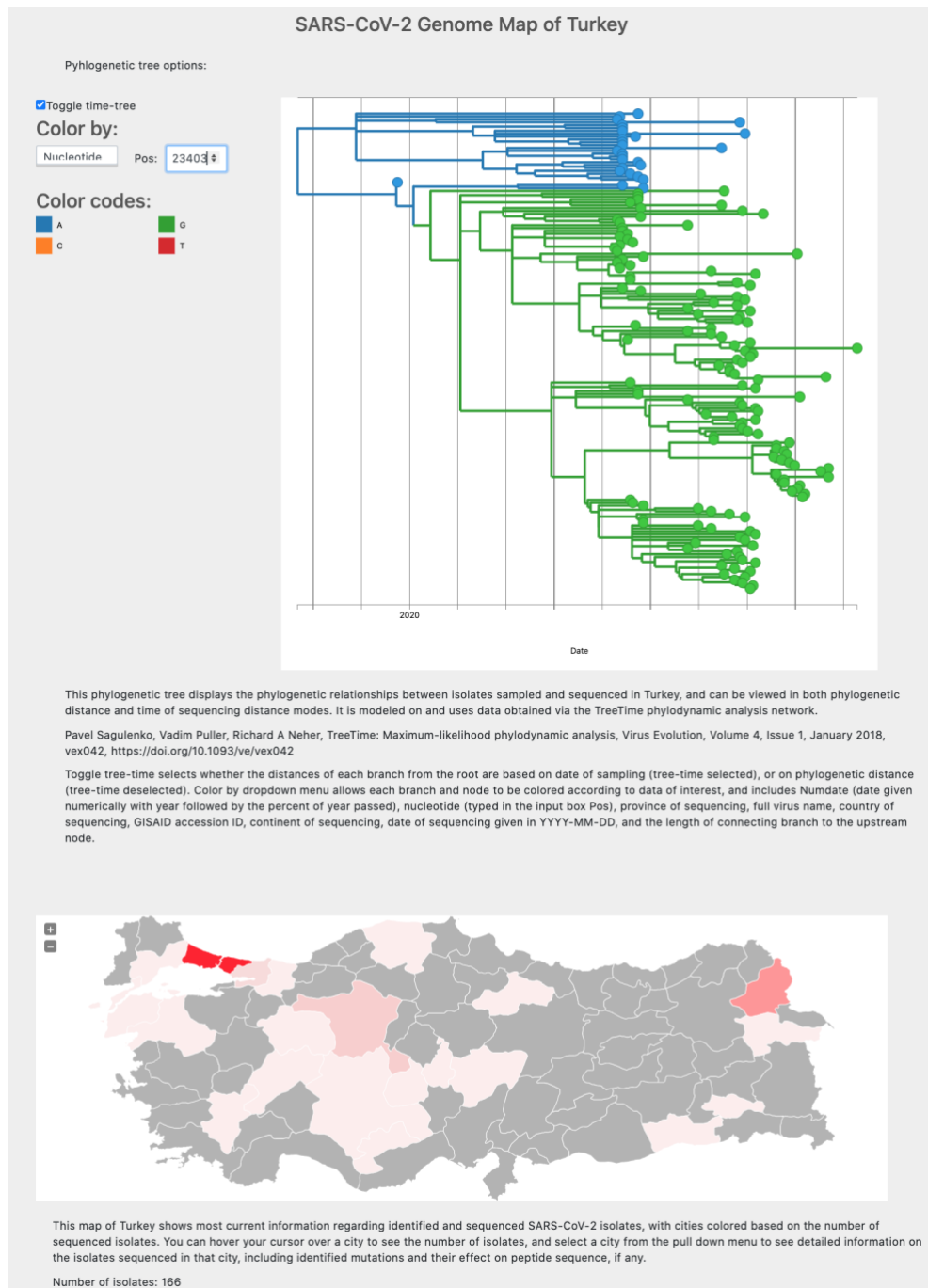
12 Tylor S, Andonov A, Cutts T, Cao J, Grudesky E, Van Domselaar G, Li X, He R (2009)
13 The SR-rich motif in SARS-CoV nucleocapsid protein is important for virus
14 replication. *Canadian Journal of Microbiology* 55: 254–260.
15 <https://doi.org/10.1139/w08-139>

16 Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic
17 variants from high-throughput sequencing data. *Nucleic Acids Research* 38:
18 e164–e164. <https://doi.org/10.1093/nar/gkq603>

19 Wong J, Jamaludin SA, Alikhan MF, Chaw L (2020) Asymptomatic transmission of
20 SARS-CoV-2 and implications for mass gatherings. *Influenza and Other*
21 *Respiratory Viruses*. <https://doi.org/10.1111/irv.12767>

- 1 Yin C (2020) Genotyping coronavirus SARS-CoV-2: methods and implications.
2 Genomics. <https://doi.org/10.1016/j.ygeno.2020.04.016>
- 3 Zhu H, Rhee J-W, Cheng P, Walianny S, Chang A, Witteles RM, Maecker H, Davis
4 MM, Nguyen PK, Wu SM (2020) Cardiovascular Complications in Patients with
5 COVID-19: Consequences of Viral Toxicities and Host Immune Response.
6 Current Cardiology Reports 22: 32. <https://doi.org/10.1007/s11886-020-01292-3>
- 7
- 8

1 **Figure**



2

3 **Figure. Snapshot of SARS-CoV-2 Genome Map of Turkey database.** Due to size

4 constraints, tables showing information on individual isolates, or summaries of

5 individual variants, are not included.

6

1 **Table 1.** Recurring mutations in Turkey.

Position	Reference	Variant	Frequency	Cities	Exonic Function	Gene	Aminoacid Change
23403	A	G	140	İstanbul(65), Karaman(1), Kastamonu(1), Nevşehir(2), Ankara(13), Kocaeli(5), Siirt(1), Aksaray(1), Sakarya(3), Afyon(1), Balıkesir(1), Konya(1), Denizli(2), Tekirdağ(1), Tokat(1), Kars(41)	nonsynonymous SNV	S	p.D614G
3037	C	T	139	İstanbul(65), Karaman(1), Kastamonu(1), Nevşehir(2), Ankara(13), Kocaeli(4), Siirt(1), Aksaray(1), Sakarya(3), Afyon(1), Balıkesir(1), Konya(1), Denizli(2), Tekirdağ(1), Tokat(1), Kars(41)	synonymous SNV	ORF1a;ORF1ab	p.F924F
14408	C	T	139	İstanbul(65), Karaman(1), Kastamonu(1), Nevşehir(2), Ankara(13), Kocaeli(4), Siirt(1), Aksaray(1), Sakarya(3), Afyon(1), Balıkesir(1), Konya(1), Denizli(2), Tekirdağ(1), Tokat(1), Kars(41)	nonsynonymous SNV	ORF1ab	p.P4715L
241	C	T	121	İstanbul(66), Nevşehir(1), Kocaeli(3), Ankara(6), Denizli(1), Ađr(1), Tekirdağ(1), Tokat(1), Kars(41)		ORF1a;ORF1ab	
28881	G	A	73	İstanbul(40), Sakarya(2), Kocaeli(3), Ankara(3), Kars(25)	nonsynonymous SNV	ORF9	p.R203K
28882	G	A	73	İstanbul(40), Sakarya(2), Kocaeli(3), Ankara(3), Kars(25)	synonymous SNV	ORF9	p.R203R
28883	G	C	73	İstanbul(40), Sakarya(2), Kocaeli(3), Ankara(3), Kars(25)	nonsynonymous SNV	ORF9	p.G204R
25563	G	T	61	İstanbul(24), Karaman(1), Kastamonu(1), Nevşehir(2), Ankara(8), Kocaeli(1), Siirt(1), Aksaray(1), Sakarya(1), Afyon(1), Balıkesir(1), Konya(1), Tekirdağ(1), Tokat(1), Kars(16)	nonsynonymous SNV	ORF3a	p.Q57H
18877	C	T	58	İstanbul(24), Kastamonu(1), Nevşehir(2), Ankara(7), Kocaeli(1), Siirt(1), Aksaray(1), Sakarya(1), Afyon(1), Balıkesir(1), Konya(1), Tekirdağ(1), Tokat(1), Kars(15)	synonymous SNV	ORF1ab	p.L6205L
7765	C	T	35	İstanbul(18), Siirt(1), Nevşehir(1), Tekirdağ(1), Kars(14)	synonymous SNV	ORF1a;ORF1ab	p.S2500S
17690	C	T	35	İstanbul(18), Siirt(1), Nevşehir(1), Tekirdağ(1), Kars(14)	nonsynonymous SNV	ORF1ab	p.S5809L

11083	G	T	31	Kayseri(1), İstanbul(11), Karaman(1), Ankara(6), Balıkesir(1), Çanakkale(1), Eskişehir(1), Kocaeli(4), Mardin(1), Ağrı(1), Kars(3)	nonsynonymous SNV	ORF1a;ORF1ab	p.L3606F
29742	G	T	24	Kayseri(1), Ankara(5), Kocaeli(5), Balıkesir(1), Çanakkale(1), Eskişehir(1), Mardin(1), İstanbul(8), Ağrı(1)		ORF10;ORF9	
1397	G	A	23	Kayseri(1), Ankara(5), Balıkesir(1), Çanakkale(1), Eskişehir(1), Kocaeli(4), Mardin(1), İstanbul(8), Ağrı(1)	nonsynonymous SNV	ORF1a;ORF1ab	p.V378I
12809	C	T	23	İstanbul(20), Kars(3)	nonsynonymous SNV	ORF1a;ORF1ab	p.L4182F
28688	T	C	23	Kayseri(1), Ankara(5), Balıkesir(1), Çanakkale(1), Eskişehir(1), Kocaeli(4), Mardin(1), İstanbul(8), Ağrı(1)	synonymous SNV	ORF9	p.L139L
27703	G	T	20	Kars(20)	nonsynonymous SNV	ORF7a	p.V104F
24262	G	T	20	Kars(20)	nonsynonymous SNV	S	p.M900I
313	C	T	19	Kars(19)	synonymous SNV	ORF1a;ORF1ab	p.L16L
2509	C	T	19	Kars(19)	synonymous SNV	ORF1a;ORF1ab	p.P748P
13620	C	T	19	Kars(19)	synonymous SNV	ORF1ab	p.D4452D
14724	C	T	19	Kars(19)	synonymous SNV	ORF1ab	p.F4820F
19839	T	C	18	Ankara(2), İstanbul(15), Kars(1)	synonymous SNV	ORF1ab	p.N6525N
26735	C	T	15	Kastamonu(1), Nevşehir(1), Ankara(5), Kocaeli(1), Aksaray(1), Sakarya(1), Afyon(1), Balıkesir(1), Konya(1), İstanbul(2)	synonymous SNV	ORF5	p.Y71Y

8326	C	T	14	Kars(14)	synonymous SNV	ORF1a;ORF1ab	p.D2687D
2113	C	T	13	İstanbul(11), Nevşehir(1), Tekirdağ(1)	synonymous SNV	ORF1a;ORF1ab	p.I616I
884	C	T	12	Balıkesir(1), Çanakkale(1), Eskişehir(1), Kocaeli(3), Ankara(3), İstanbul(3)	nonsynonymous SNV	ORF1a;ORF1ab	p.R207C
8653	G	T	12	Balıkesir(1), Çanakkale(1), Eskişehir(1), Kocaeli(3), Ankara(3), İstanbul(3)	nonsynonymous SNV	ORF1a;ORF1ab	p.M2796I
26549	C	T	12	Kocaeli(2), İstanbul(5), Ankara(3), Tokat(1), Denizli(1)	synonymous SNV	ORF5	p.T9T
5015	G	A	11	Kars(11)	nonsynonymous SNV	ORF1a;ORF1ab	p.V1584M
28854	C	T	10	Ankara(3), Aksaray(1), Sakarya(1), Konya(1), İstanbul(3), Kars(1)	nonsynonymous SNV	ORF9	p.S194L
228	C	T	9	Kocaeli(1), İstanbul(4), Ankara(3), Denizli(1)		ORF1a;ORF1ab	
22444	C	T	9	Ankara(3), Aksaray(1), Sakarya(1), Konya(1), İstanbul(3)	synonymous SNV	S	p.D294D
9514	A	G	8	Kocaeli(1), İstanbul(4), Ağrı(1), Ankara(2)	synonymous SNV	ORF1a;ORF1ab	p.L3083L
26720	G	C	8	Kocaeli(1), İstanbul(4), Ağrı(1), Ankara(2)	synonymous SNV	ORF5	p.V66V
9479	G	T	7	Kocaeli(1), İstanbul(4), Ağrı(1), Ankara(1)	nonsynonymous SNV	ORF1a;ORF1ab	p.G3072C
28835	T	C	7	Kocaeli(1), İstanbul(4), Ağrı(1), Ankara(1)	nonsynonymous SNV	ORF9	p.S188P

5736	C	T	6	Ankara(3), Denizli(1), İstanbul(2)	nonsynonymous SNV	ORF1a;ORF1ab	p.A1824V
16428	C	T	6	Kars(6)	synonymous SNV	ORF1ab	p.Y5388Y
25611	C	A	6	İstanbul(6)	synonymous SNV	ORF3a	p.L73L
28857	G	T	6	Kars(6)	nonsynonymous SNV	ORF9	p.R195I
10702	C	T	5	Eskişehir(1), Ankara(1), İstanbul(3)	synonymous SNV	ORF1a;ORF1ab	p.D3479D
20268	A	G	5	Ankara(2), Denizli(1), Kocaeli(1), İstanbul(1)	synonymous SNV	ORF1ab	p.L6668L

1

1 **Table 2.** Frequencies of SARS-CoV-2 mutations in different geographical regions

Mutation	Africa	Asia	Europe	North America	Oceania	South America	Turkey	Worldwide
23403A>G	88.73% (252)	58.15% (835)	82.34% (10484)	78.15% (4470)	67.07% (894)	93.48% (258)	84.34% (140)	78.94% (17193)
3037C>T	79.58% (226)	58.84% (845)	82.03% (10444)	78.25% (4476)	66.99% (893)	93.48% (258)	83.73% (139)	78.7% (17142)
14408C>T	87.68% (249)	58.57% (841)	82.32% (10481)	78.36% (4482)	67.14% (895)	93.48% (258)	83.73% (139)	79% (17206)
241C>T	88.73% (252)	58.29% (837)	82.24% (10471)	77.36% (4425)	51.24% (683)	93.48% (258)	72.89% (121)	77.71% (16926)
28881G>A	22.18% (63)	16.92% (243)	38.73% (4931)	5.03% (288)	13.5% (180)	48.55% (134)	43.98% (73)	26.81% (5839)
28882G>A	22.18% (63)	16.85% (242)	38.68% (4925)	5.02% (287)	13.28% (177)	48.55% (134)	43.98% (73)	26.76% (5828)
28883G>C	22.18% (63)	16.92% (243)	38.67% (4924)	5.02% (287)	13.35% (178)	48.55% (134)	43.98% (73)	26.76% (5829)
25563G>T	10.21% (29)	26.46% (380)	12.21% (1554)	65.93% (3771)	28.21% (376)	30.8% (85)	36.75% (61)	28.44% (6195)
18877C>T	2.46% (7)	15.67% (225)	1.2% (153)	5.38% (308)	1.73% (23)	9.78% (27)	34.94% (58)	3.41% (743)
7765C>T	1.06% (3)	0.35% (5)	0.57% (73)	0.07% (4)	0.15% (2)	0% (0)	21.08% (35)	0.4% (87)
17690C>T	1.06% (3)	0.14% (2)	0.47% (60)	0.07% (4)	0.08% (1)	0% (0)	21.08% (35)	0.32% (70)
11083G>T	8.8% (25)	28.41% (408)	10.91% (1389)	3.22% (184)	15.68% (209)	5.07% (14)	18.67% (31)	10.23% (2229)
29742G>T	0.7% (2)	2.44% (35)	0.13% (17)	0.24% (14)	3.45% (46)	0% (0)	14.46% (24)	0.52% (114)
1397G>A	1.06% (3)	2.44% (35)	0.03% (4)	0.21% (12)	4.43% (59)	0% (0)	13.86% (23)	0.52% (113)

12809C>T	0% (0)	0% (0)	0.02% (3)	0.12% (7)	0% (0)	0% (0)	13.86% (23)	0.05% (10)
28688T>C	1.06% (3)	2.51% (36)	0.02% (3)	0.14% (8)	4.35% (58)	0% (0)	13.86% (23)	0.5% (108)
27703G>T	0% (0)	0% (0)	0.06% (8)	0% (0)	0.08% (1)	0% (0)	12.05% (20)	0.04% (9)
313C>T	0.7% (2)	4.81% (69)	1.56% (199)	0.72% (41)	0.75% (10)	0.36% (1)	11.45% (19)	1.48% (322)
13620C>T	0% (0)	0% (0)	0.04% (5)	0.03% (2)	0% (0)	0% (0)	11.45% (19)	0.03% (7)
14724C>T	0% (0)	0% (0)	0.07% (9)	0.12% (7)	0.08% (1)	0.36% (1)	11.45% (19)	0.08% (18)
19839T>C	0.7% (2)	0.63% (9)	2.37% (302)	0.59% (34)	0.38% (5)	0.36% (1)	10.84% (18)	1.62% (353)
26735C>T	0.35% (1)	14.35% (206)	0.57% (73)	0% (0)	0.6% (8)	0% (0)	9.04% (15)	1.32% (288)
8326C>T	0% (0)	0.07% (1)	0.03% (4)	0% (0)	0% (0)	0% (0)	8.43% (14)	0.02% (5)
2113C>T	0.7% (2)	0.14% (2)	0.49% (62)	0.03% (2)	0.08% (1)	0% (0)	7.83% (13)	0.32% (69)
884C>T	0.7% (2)	1.95% (28)	0.02% (3)	0.1% (6)	0.38% (5)	0% (0)	7.23% (12)	0.2% (44)
8653G>T	0.7% (2)	1.88% (27)	0.04% (5)	0.12% (7)	0.3% (4)	0% (0)	7.23% (12)	0.21% (45)
28854C>T	0.7% (2)	6.69% (96)	2.52% (321)	1.38% (79)	0.83% (11)	0.36% (1)	6.02% (10)	2.34% (510)
228C>T	0.35% (1)	0.07% (1)	0.02% (3)	0% (0)	0.08% (1)	0% (0)	5.42% (9)	0.03% (6)
22444C>T	0% (0)	6.48% (93)	0.02% (2)	0% (0)	0% (0)	0% (0)	5.42% (9)	0.44% (95)
9479G>T	0.35% (1)	0.14% (2)	0.07% (9)	0.05% (3)	0.15% (2)	0% (0)	4.22% (7)	0.08% (17)
16428C>T	0% (0)	0% (0)	0.01% (1)	0.12% (7)	0% (0)	0.36% (1)	3.61% (6)	0.04% (9)
28857G>T	0% (0)	0.07% (1)	0.09% (12)	0.02% (1)	0% (0)	0% (0)	3.61% (6)	0.06% (14)

20268A>G	3.17% (9)	1.04% (15)	8.18% (1042)	1.03% (59)	4.13% (55)	9.06% (25)	3.01% (5)	5.53% (1205)
----------	-----------	------------	--------------	------------	------------	------------	-----------	--------------

1

1 **Table 3.** Mean number of variants per isolate in different geographical regions.

Region	Mean Number of Variants per Isolate
Turkey	10.18
Worldwide	8.01
Africa	8.47
Asia	8.53
Europe	7.88
North America	8.20
Oceania	7.54
South America	9.32

2

1 **Table 4.** SNV densities of SARS-CoV-2 genes in different geographical regions.

Region	Orf1ab	S	Orf3	E	M	Orf6	Orf7a	Orf7b	Orf8	N	Orf10
Turkey	0.25	0.32	0.49	0.06	0.38	0.03	0.44	0.05	0.11	1.45	0.06
Worldwide	0.19	0.29	0.62	0.08	0.13	0.11	0.11	0.10	0.38	0.84	0.10
Africa	0.21	0.40	0.25	0.06	0.15	0.00	0.06	0.00	0.23	0.75	0.12
Asia	0.20	0.31	0.48	0.11	0.31	0.10	0.13	0.12	0.36	0.83	0.10
Europe	0.18	0.29	0.48	0.09	0.13	0.09	0.12	0.09	0.15	1.10	0.11
North America	0.20	0.26	0.97	0.06	0.09	0.12	0.08	0.12	0.84	0.27	0.11
Oceania	0.19	0.26	0.66	0.05	0.13	0.08	0.07	0.05	0.62	0.75	0.06
South America	0.16	0.33	0.59	0.08	0.06	1.68	0.11	0.14	0.11	1.46	0.06