

A cross-sectional overview of SARS-CoV-2 genome variations in Turkey

Koray Ergünay

Hacettepe University: Hacettepe Universitesi

Mücahit Kaya

Ondokuz Mayıs University: Ondokuz Mayıs Universitesi

Muhittin Serdar

Acıbadem Mehmet Ali Aydınlar Üniversitesi: Acıbadem Universitesi

Yakut Akyön

Hacettepe University: Hacettepe Universitesi

Engin Yılmaz (✉ eyilmaz@hacettepe.edu.tr)

Hacettepe Universitesi Tıp Fakültesi <https://orcid.org/0000-0001-8873-7645>

Research Article

Keywords: COVID-19, SARS-CoV-2, genome, mutation

Posted Date: April 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-472330/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Introduction: Nearly a year following the emergence of COVID-19 in Turkey, we analysed SARS-CoV-2 sequences to identify virus genome variations and their probable impact on epidemiology, immune response and clinical disease.

Materials and Methods: Complete genomes and partial Spike (S) region sequences originating from Turkey were accessed from the Global Initiative on Sharing Avian Influenza Data (GISAID) database. The genomes were aligned and analysed for variations and recombinations using appropriate softwares.

Results: 410 complete genomes and 206 S region sequences were included. Overall, 1200 distinct nucleotide variations were noted. Mean variation count was noted as 14.2 per genome and increased significantly during the course of the pandemic. The most frequent variations were identified as A23403G (D614G; 92.9%), C14408T (P323L, 92.2%), C3037T (89.8%), C241T (83.4%) and GGG28881AAC (RG203KR, 62.6%). The A23403G mutation was the most frequent variation in the S region sequences (99%). Majority of the genomes (%98.3) belonged in the SARS-CoV-2 haplogroup A. No evidence for recombination was identified in genomes representing sub-haplogroup branches. The variants of concern B.1.1.7, B.1.351 and P.1 were detected, with a statistically-significant time-associated increase in the variant B.1.1.7 prevalence.

Discussion: We described prominent SARS-CoV-2 variations as well as comparisons with global virus diversity. Continuing a molecular surveillance in agreement with local disease epidemiology appears to be crucial, as vaccination and mitigation efforts are ongoing.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the pandemic of the twentyfirst century, also called as Coronavirus Disease-2019 (COVID-19) [1,2]. Initially emerging as cases of atypical viral pneumonia in the Wuhan city, Hubei province of China, SARS-CoV-2 have readily spread around the globe, currently affecting millions of people in more than 220 countries. More than a year after its official announcement, the control of the pandemic and mitigating its health, economic and social impact have become the foremost priority in all countries and international organizations.

SARS-CoV-2, classified in the *Betacoronavirus* genus of the family *Coronaviridae*, is an enveloped particle with a positive-stranded RNA as its genome [3]. The viral genome is approximately 29.9 kb in size and flanked by untranslated regions (UTRs) in the 3' and 5' ends. Virus proteins comprising the structural components nucleocapsid (N), envelope (E), membrane (M) and spike (S), as well as several accessory and non-structural proteins (NSPs) are encoded by several open reading frames (ORFs) on the genome [4,5]. Variations occur spontaneously during SARS-CoV-2 replication but mutation frequency is relatively low compared to other RNA viruses, owing to the activity of nsp14 exoribonuclease [6,7]

The exploration of SARS-CoV-2 genetic diversity in has been pivotal for control efforts since the beginning of the pandemic. Sequencing of the virus genome has provided critical information for the investigation of origin and spread of the infection as well as development of appropriate diagnostics. It further enables surveillance of virus molecular epidemiology and identification of variants that can undermine mitigation and vaccination efforts [8,9]. Aided by the widespread availability of next generation sequencing technologies and online sharing of genome information, a significant number of virus genomes originating from several countries affected by the pandemic has now been available. As the vaccination efforts are spreading in many countries, the monitorization of the virus genetic diversity will continue to be crucial to identify variants with enhanced transmissibility or clinical disease as well as potential escape mutants [9]. In this study, we aimed to evaluate SARS-CoV-2 genomic diversity from Turkey since the emergence of initial cases, to gain insights into origin, patterns of variation and their probable impact on epidemiology.

Materials And Methods

The Global Initiative on Sharing Avian Influenza Data" (GISAID) [10] database (<https://www.gisaid.org>) was screened for SARS-CoV-2 sequences originating from Turkey and all complete virus genomes dated between 16.03.2020 and 30.01.2021 were accessed (**Supplement 1**). High quality virus genomes larger than 29.000 nucleotides with less than 1% N content and 0.05% specific amino acid variation were selected. Recurrent submissions and genomes with unverified insertion/deletions were omitted. The SARS-CoV-2 Wuhan-Hu-1 isolate genome sequence (GISAID access: EPI_ISL_402125, GenBank access: MN908947.3, NC045512) was used as the reference. Sequences were aligned using MUSCLE within the MSA package (v1.22.0), based on the R platform (v4.0.3, 2020-10-10) [11,12]. Nucleotide variations were identified by the Sequence Variation (SNP) package in the Virus Pathogen Resource (ViPR) database and further confirmed via MEGAX [13-15]. Probable recombinations were screened using the R8DP, Geneconv, Bootscan, MaxChi, Chimaera, SiScan, 3Seq tools of the RDP4 software in the default settings [16]. A total of 410 genomes were analysed.

To incorporate recent data in the analysis, we further screened spike region sequences uploaded at a later date in the database and accessed a total of 206 non-redundant sequences, collected between 30.04.2020-03.02.2021 but mostly during January 2021 (170/206) (**Supplement 1**). Descriptive statistics, data distribution and correlations were analysed using Analyse-it v4.20.1 (Analyse-it Software Ltd. Leeds, United Kingdom).

Results

Among the 410 SARS-CoV-2 genomes included in the study, 261 (63%) originated from specimens collected during March-June 2020, 27 (6.58%) in July-October 2020 and 122 (29.75%) in December 2020 - January 2021. Age, gender, residence and other demographic or clinical features of infected cases were not evaluated due to the lack of sufficient information in the database.

Complete genome sequencing revealed a total of 1200 different nucleotide variations (**Table 1**). Majority of the variations were missense (661/1200, 55.1%) and silent mutations (441/1200, 36.7%). The variations were frequently located in ORF1a/1b region (684/1200, 57%), that encodes for the proteins and cofactors participating in viral replication; as well as in S region (153/1200, 12.5%), involved in virus attachment on the respiratory epithelium and the main target for neutralizing antibodies (**Table 1**).

Median variation count per genome was calculated as 12 (mean+standard deviation: 14.2+6.5, range: 4-36). Temporal distribution of the total variation count per genome according to sampling date revealed a positive correlation and statistically-significant increase during the course of the pandemic (**Figure 1**).

Variations frequently-identified in the SARS-CoV-2 genomes were overviewed in **Figure 2**. Here, the most commonly-observed variations representing 62.2% - 92.9% of the diversity repertoire resulted in amino acid substitution; whereas silent changes and non-ORF mutations were also noted (**Table 2**). Among these, the A23403G, C14408T and GGG28881AAC variations stood out, as they are reported to be associated with severe infections requiring hospitalization [8]. Further screening for previously-reported globally-prevalent variations revealed G25563T (ORF3, Q57H) (n=92), C1059T (nsp2, L37F) (n=9), C8782T (n=7) and C17747T (nsp13, P504L) (n=1) in the study groups [8,17].

We also checked particular mutations, previously-reported to potentially affect predicted T and B cell epitopes in S, N ve M proteins [18]. The most frequent of such mutations is the A23403G, resulting in the D614G substitution, which is also reported to enhance virus infectivity as well as B cell epitopes in S1 spike region [18,19] (**Table 2**). Other include the S region T cell epitope variations C24378T (n=2) ve C24381T (n=2), and N region B cell epitope variations G29383A (n=1) and G29405C (n=2)

We further evaluated the diversity of the SARS-CoV-2 genomes as previously-described mutation-annotated virus lineages and haplotypes [8]. Majority of the genomes (403/410, 98.3%) were observed to belong in haplogroup A, from which various subhaplogroups were further identified (A2: n=381, 92.9%; A3: n=39, 9.5%; A9: n=29, 7.1%; A12: n=2, 0.5%) (**Figure 3**). The most frequent subhaplogroup was noted as A2a, that diversifies from A2 by the C14408T variation. Haplogroup B genomes were relatively few (7/410, 1.7%), with subhaplogroups B4 and B4a represented with 8 and 23 genomes, respectively (**Figure 3**). No evidence of recombination was observed among viruses of the A2a2b2a, A3a1a, A12 and B4a subhaplogroups, representing highly-diversified genomes within the group.

In the spike region dataset, a total of 44 nucleotide variations comprising missense (32/206, 72.7%) and silent (12/206, 29.3%) mutations (**Figure 4**). The most frequent variation was observed as A23403G (204/206, 99%), similar to the complete genome dataset. Of the virus epitopes located in the spike protein, the C24378T (S939F) variation, affecting predicted T cell epitope was detected in 2 (0.9%) sequences.

Finally, particular SARS-CoV-2 variants with potential impact on transmission, disease course and immune protection were screened in the study groups [20-22]. These were detected in 30.2% (186/616) of the individual viruses where the B.1.1.7 variant (variant 1 or 501Y.V 1) was noted as the most frequent

(176/186, 94.6%) (**Table 3**). In the complete genomes, B.1.1.7 and B.1.351 (501Y.variant 2) were relatively scarce, but was significantly-increased in the spike region dataset (chi-square test, $p < 0.001$), where B 1.1.28 (variant P.1) also became detectable (**Table 3**)

Discussion

In this study, we provide a cross-sectional overview and potential impact of the SARS-CoV-2 genome variations from Turkey. The genomes were accessed from the GISAID database, originated from specimens collected during a 10-month period in 2020 and 2021. We further included additional S region sequences, submitted after our initial database searches as an update, representing viruses circulating in early 2021. Therefore, the findings are based on these datasets of 410 complete and 206 partial (spike) sequences, being so far the most comprehensive analysis performed in Turkey [23-25].

In the complete SARS-CoV-2 genomes, we identified 1200 individual nucleotide variations, with a median frequency of 12 (range:4-36) per genome (**Table 1**). Moreover, the temporal distribution of the variations indicated a statistically-significant accumulation of variations during the 10-month period examined (**Figure 1**). Comparable findings were reported in globally-distributed virus isolates, with more than 3000 specific point mutations being detected and an increased frequency of variation during the course of the pandemic [8]. However, SARS-CoV-2 isolates from Turkey was proposed to exhibit an elevated variation rate in a study focusing on 166 virus genomes accessed during July 2020 [25], where frequently-detected variations C14408T and C18877T, affecting viral polymerase (nsp12) and exoribonuclease (nsp14), respectively; were suggested as a possible precipitating factors [27,28]. These variations were also noted in our study with varying rates (**Table 2, Figure 2**). In addition, the co-detection of C14408T and A23403G variations were suggested to be associated with increased diversity [26,27] In parallel with global isolates, the SARS-CoV-2 genome variations in Turkey are mostly missense or silent mutations, frequently involving the enzymes and co-factors, participating in replication in ORF1a/1b or the S regions of the virus genome [8,17,26].

In the study, the most frequently-detected variations, namely the A23403G, C14408T and GGG28881AAC mutations resulting in amino acid substitutions in the corresponding virus proteins, were reported in previous analyses from Turkey [23-25]. However, they seem to be positively-selected in the local virus population pool, as their abundance seem to be elevated. For example, the A23403G variation was reported as low as 56.2% in previous reports, while it is detected in 92.9% of the complete genomes and 99% in S regions in this study (**Figure 2, Figure 4**). This observation is also evident in global genome data, where viruses with the A23403G and C14408T variations were steadily increased in frequency during the course of the pandemic and have become the majority in late 2020 [8]. The amino acid substitutions occurring as a result of these variations, namely P323L, D614G ve RG203KR, were also associated with a more severe COVID-19 clinical presentation [8]. Moreover, the D614G mutation, a defining component of the variant of concern B.1.1.7, is also likely to affect immune responses to the S protein. In addition to the high frequency of this substitution in the study, we further identified other variations that might affect T and B cell epitopes, albeit with lower rates.

Throughout the pandemic, the availability of virus genome sequences and powerful online tools have enabled a nearly real-time monitorization of SARS-CoV-2 molecular epidemiology [8,17]. Previous reports on SARS-CoV-2 genetic diversity have described particular virus lineages and clades, mostly in overall agreement but lacking a uniform nomenclature [8,29]. The size of accumulating sequence data further warrants more practical approaches to indicate phylogeographic relationships than standard phylogenetic reconstruction. Here, we adopted a previously-reported mutation-annotated reference strategy to describe intraspecific phylogeny of SARS-CoV-2. We observed the majority of the SARS-CoV-2 genomes from Turkey to belong in the haplogroup A (98.3%), with main subhaplotype diversification into A2 (**Figure 3**). SARS-CoV-2 haplogroup A isolates constitute the ancestral node and predominant clade across the world. They are frequently-represented in isolates from Europe (97%) Africa (93%) and Asia (77%), but relatively scarce in South America (68%) and North America (53%) [8]. Among global haplogroup A subclades, A2 and A2a appears as the majority, with the phylogeographic inferences indicating a European origin. We observed haplogroup B is with a much lower frequency, and B4a subhaplotype representing the majority within this group (**Figure 3**). Haplogroup B viruses have been identified in all continents, with higher prevalence in North America (47%), South America (32%), Asia and Oceania (23%), with all major and minor subclades present in Asia [8]. The haplotype B viruses were introduced in Turkey likely by travel to endemic regions and further local spread was presumably prevented by isolation. Our analyses employing highly-diversified subclades of each main haplogroup failed to identify any evidence for recombination among local SARS-CoV-2 genomes. Overall, the findings on virus genome diversity in Turkey suggest several introductions originating from multiple sources and subsequent local adaptation, also noted in previous reports using smaller datasets [23].

The emergence and rapid spread of SARS-CoV-2 variants has raised significant concern, due to their potential for enhanced transmissibility, altered clinical progression and escape from protective immune response induced by previous infection or widely-available vaccines [20]. Also dubbed as the variant of concern (VOC), these viruses exhibit a wide array of amino acid changes accumulated in several regions of the virus genome including the spike protein [20-22]. The rapid spread of particular VOCs in several countries during fall 2020 called for more stringent public health measures as well as targeted monitorization, which is also initiated and currently ongoing in Turkey. We detected three major VOCs in the study group, with increased prevalence of B.1.1.7 and B.1.351 in the recently-dated dataset (**Table 3**). Moreover, the detection of P.1 in this group suggests not only an elevated prevalence but also a broader repertoire of variants in the population. These findings justify the efforts to identify and monitor known and potentially-emerging virus variants.

Particular limitations of this study need to be addressed. An important issue is the heterogeneity in temporal and spatial distribution of the samples employed for genome sequencing, which suggests a lack of organized sampling strategy for screening. In addition, missing demographic and location data in many instances also prevented further evaluations. Therefore, it is not possible to assess whether the current dataset fully represents the epidemiology and diversity in circulating viruses in Turkey. A continuous and organized surveillance strategy in conjunction with local transmission dynamics and

infection epidemiology, will provide a better understanding of the SARS-CoV-2 molecular epidemiology in Turkey.

In conclusion, in this analysis of complete and partial SARS-CoV-2 genome sequences almost covering the first year since emergence, we described main variations associated with epidemiology and immune response, with the observation of increased incidence of VOCs in Turkey. With the ongoing pandemic and accelerated vaccination campaigns, such investigations should be performed periodically for precise screening and coordination of control measures.

Declarations

Funding: No funding was received for the study

Conflicts of interest/Competing interests: The authors have no conflicts of interest to declare with third parties

Availability of data and material (data transparency): All data used in the manuscript are available online or provided in the submission

Ethics approval: The study and its findings are based on virus genome sequences obtained from online sources, therefore, no institutional ethics board approval was required nor sought. The study was approved by the COVID-19 Scientific Research Evaluation Committee, the Ministry of Health of Turkey (2021-03-22T11_13_48).

Acknowledgement : The authors are grateful to Diagen Biyoteknolojik Sistemler Inc., Turkey, for technical support during sequence handling. Information on GISAID sequence records is provided as a separate file (**Acknowledgement**).

References

1. Zhu N, Zhang D, Wang W et al (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382:727–733
2. Cucinotta D, Maurizio Vanelli M (2020). WHO Declares COVID-19 a Pandemic. *Acta Biomed* 91:157–160
3. Gorbalenya AE, Baker SC, Baric RS et al (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 5:536–544
4. Lu R, Zhao X, Li J et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395:565–574
5. Wu A, Peng Y, Huang B et al (2020). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* 27:325–328

6. Grubaugh ND, Petrone ME, Holmes EC (2020). We shouldn't worry when a virus mutates during disease outbreaks. *Nat Microbiol* 5:529–530
7. Ogando NS, Zevenhoven-Dobbe JC, van der Meer Y, Bredenbeek PJ, Posthuma CC, Snijder EJ (2020). The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J Virol* 94:e01246–20
8. Flores-Alanis A, Cruz-Rangel A, Rodríguez-Gómez F et al (2021). Molecular epidemiology surveillance of SARS-CoV-2: mutations and genetic diversity one year after emerging. *Pathogens* 10:184
9. Gómez-Carballa A, Bello X, Pardo-Seco J, Martín-Torres F, Salas A (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Res* 30:1434–1448
10. Shu Y, McCauley J (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22:30494
11. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S (2015). MSA: an R package for multiple sequence alignment. *Bioinformatics* 31:3997–3999
12. Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
13. Pickett BE, Greer DS, Zhang Y et al (2004). Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses* 4:3209–3226
14. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004). WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190
15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
16. Martin DP, Murrell B, Khoosal A, Muhire B (2017). Detecting and analyzing genetic recombination Using RDP4. *Methods Mol Biol* 1525:433–460
17. Mercatelli D, Giorgi FM (2020). Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol* 11:1800
18. Koyama T, Weeraratne D, Snowdon JL, Parida L (2020). Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens* 9:324
19. Korber B, Fischer WM, Gnanakaran S et al (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182:812–827
20. Tang JW, Tambyah PA, Hui DS (2020). Emergence of a new SARS-CoV-2 variant in the UK. *J Infect* 82:e27–e28
21. Tegally H, Wilkinson E, Giovanetti M et al (2021). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *Nature*. <https://doi.org/10.1038/s41586-021-03402-9>

22. Toovey OTR, Harvey KN, Bird PW, Tang JWW. Introduction of Brazilian SARS-CoV-2 484K.V2 related variants into the UK. *J Infect* 82:e23–e24
23. Adebali O, Bircan A, Çirci D, İşlek B, Kiliç Z, Selçuk B, Turhan B (2020). Phylogenetic analysis of SARS-CoV-2 genomes in Turkey. *Turk J Biol* 44:146–156
24. Hanifehnezhad A, Kehribar EŞ, Öztop S et al (2020). Characterization of local SARS-CoV-2 isolates and pathogenicity in IFNAR^{-/-} mice. *Heliyon* 6(9):e05116
25. Eskier D, Akalp E, Dalan Ö, Karakulah G, Oktay Y (2021). Current mutasome of SARS-CoV-2 in Turkey reveals mutations of interest. *Turk J Biol* 45:104–113
26. Koçhan N, Eskier D, Suner A, Karakulah G, Oktay Y (2021). Different selection dynamics of S and RdRp between SARS-CoV-2 genomes with and without the dominant mutations. *Infect Genet Evol* 91:104796
27. Eskier D, Karakulah G, Suner A, Oktay Y (2020). RdRp mutations are associated with SARS-CoV-2 genome evolution. *PeerJ* 8:e9587
28. Eskier D, Suner A, Oktay Y, Karakulah G (2020). Mutations of SARS-CoV-2 nsp14 exhibit strong association with increased genome-wide mutation load. *PeerJ* 8:e10181
29. Alm E, Broberg EK, Connor T et al (2020). Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill* 25:2001410
30. Davies NG, Jarvis CI; CMMID COVID-19 Working Group, et al (2021). Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*. <https://doi.org/10.1038/s41586-021-03426-1>

Tables

Table 1: Spatial distribution of the variations observed in SARS-CoV-2 genomes.

		Variation Type					
Region	Location	Missense mutation	Silent mutation	Nonsense mutation	Deletion	Non-ORF change	<i>Total</i>
5'UTR	1-258	-	-	-	-	22	22
ORF1a	259-13476	282	107	1	-	-	390
ORF1b	13461-21548	101	193	-	-	-	294
S	21556-25377	102	51	-	-	-	153
ORF3a	25386-26213	53	17	-	-	-	70
E	26238-26465	2	2	-	-	-	4
M	26516-27184	10	19	-	-	-	29
ORF6	27195-27380	2	7	-	-	-	9
ORF7a	27387-27752	14	8	1	-	-	23
ORF7b	27749-27880	4	2	-	-	-	6
ORF8	27887-28252	18	5	48	-	-	71
N	28267-29526	68	27	1	2	5	103
ORF10	29551-29667	5	3	-	-	-	8
3'UTR	29668-29868	-	-	-	-	18	18
Total		661	441	51	2	45	1200

Table 2: Common variations and predicted outcomes in SARS-CoV-2 genomes

Variation	Frequency (<i>n</i> , %)	Target Region	Outcome
A23403G	381, 92.9	S - Spike protein	Amino acid change: D614G
C14408T	378, 92.2	Non-structural protein 12 - RNA-dependent RNA polymerase	Amino acid change: P323L
C3037T	368, 89.8	Non-structural protein 3 - phosphoesterase	Silent
C241T	342, 83.4	5'UTR	non-ORF
GGG28881AAC	255, 62.6	N - Nucleocapsid protein	Amino acid change: RG203KR

Table 3: Distribution of the current variants of concern in the study groups

Variant	Genome (n=410)		S region (n=206)		Total
	<i>n</i>	%	<i>n</i>	%	
B.1.1.7	31	7.6	145	70.4	176
B.1.351	2	0.5	7	3.4	9
P.1	-	-	1	0.5	1
<i>Total</i>	33		153	-	186

Figures

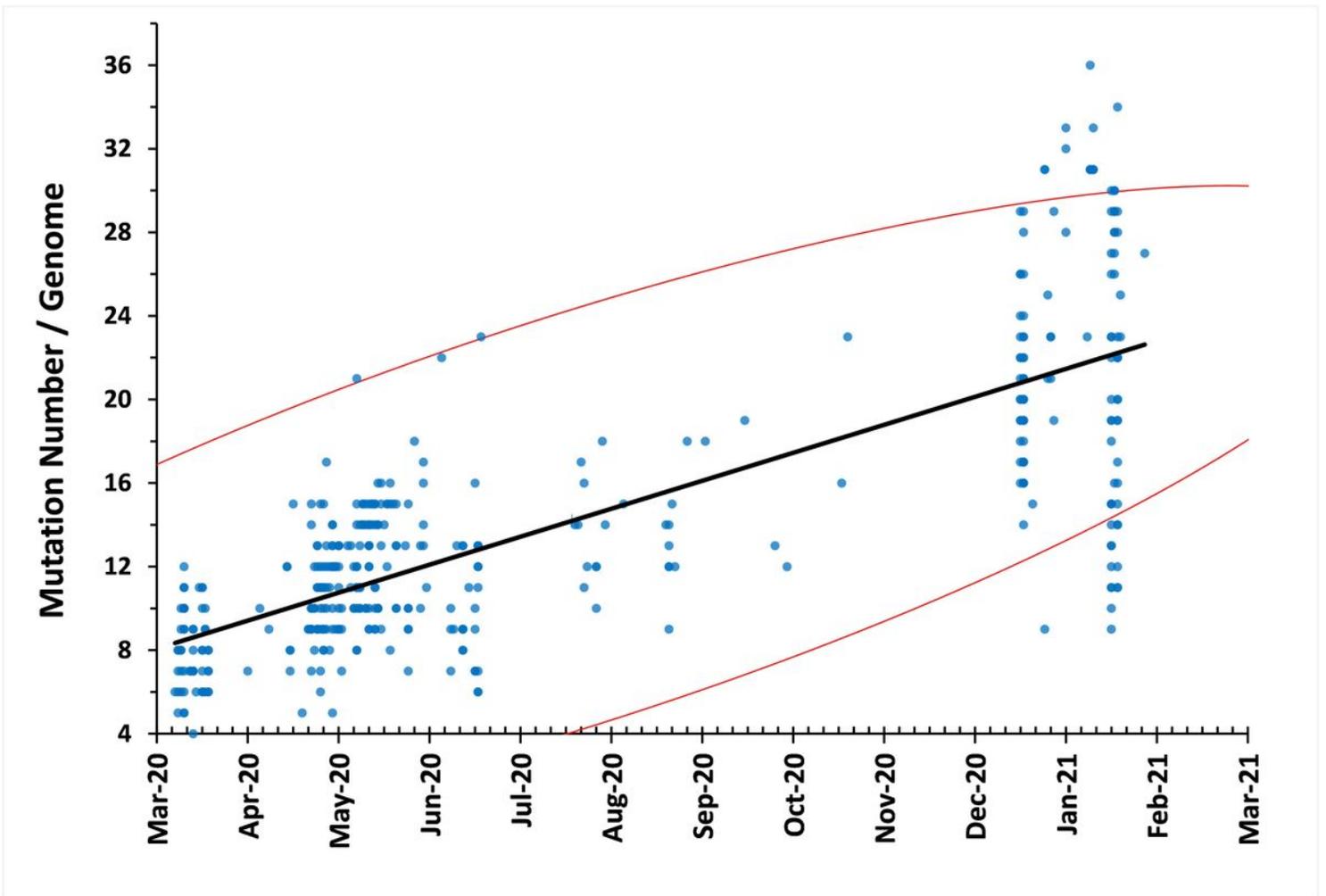


Figure 1

Temporal distribution of the total variation counts in the study. Spearman test was used to calculate correlation coefficient. Red lines indicate 95% confidence intervals. A time-related, statistically-significant increase total variation count per genome was observed ($r=0.776$, $p<0.001$).

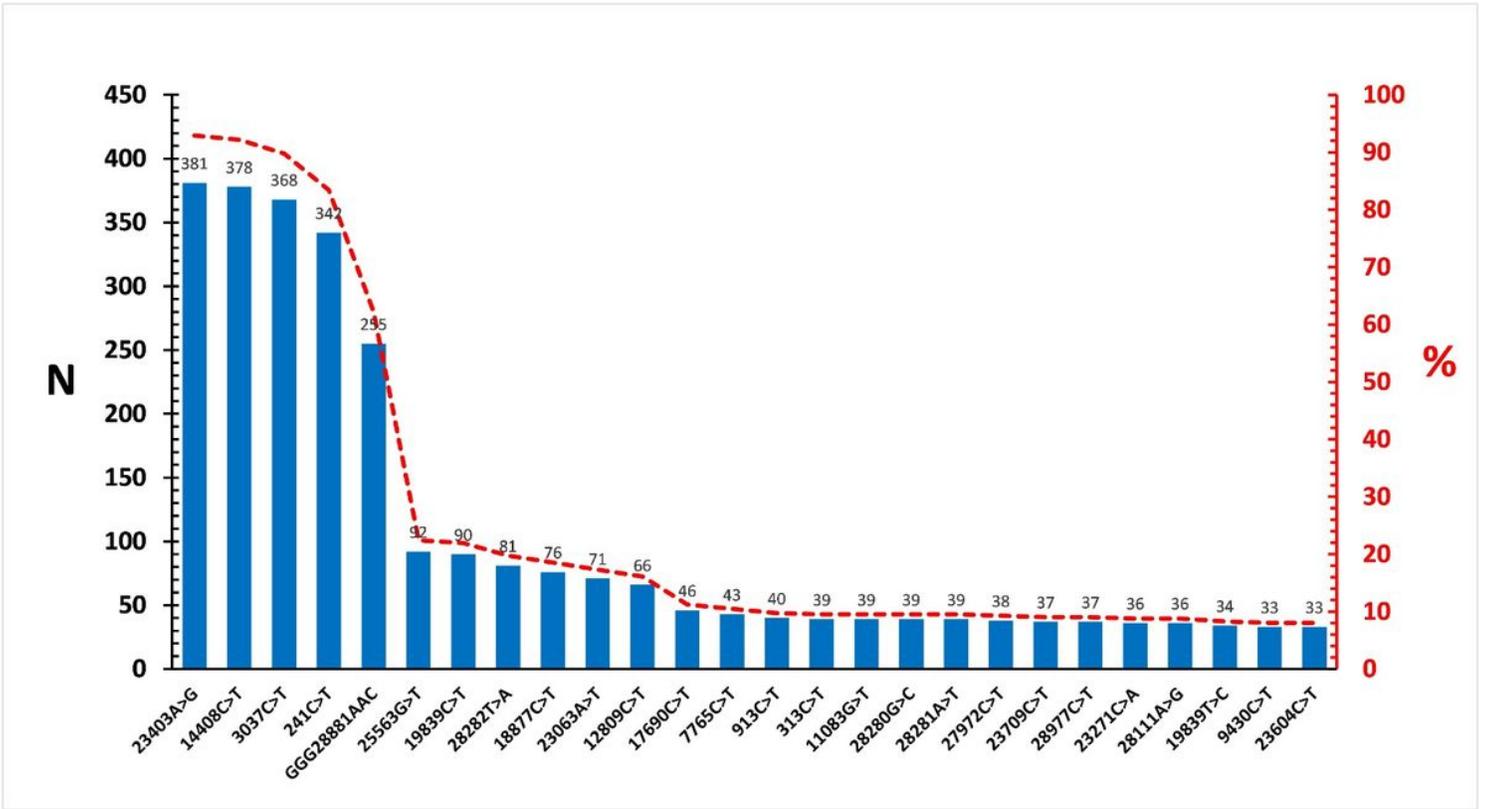


Figure 2

Distribution of the frequently-observed variations in complete SARS-CoV-2 genome dataset.

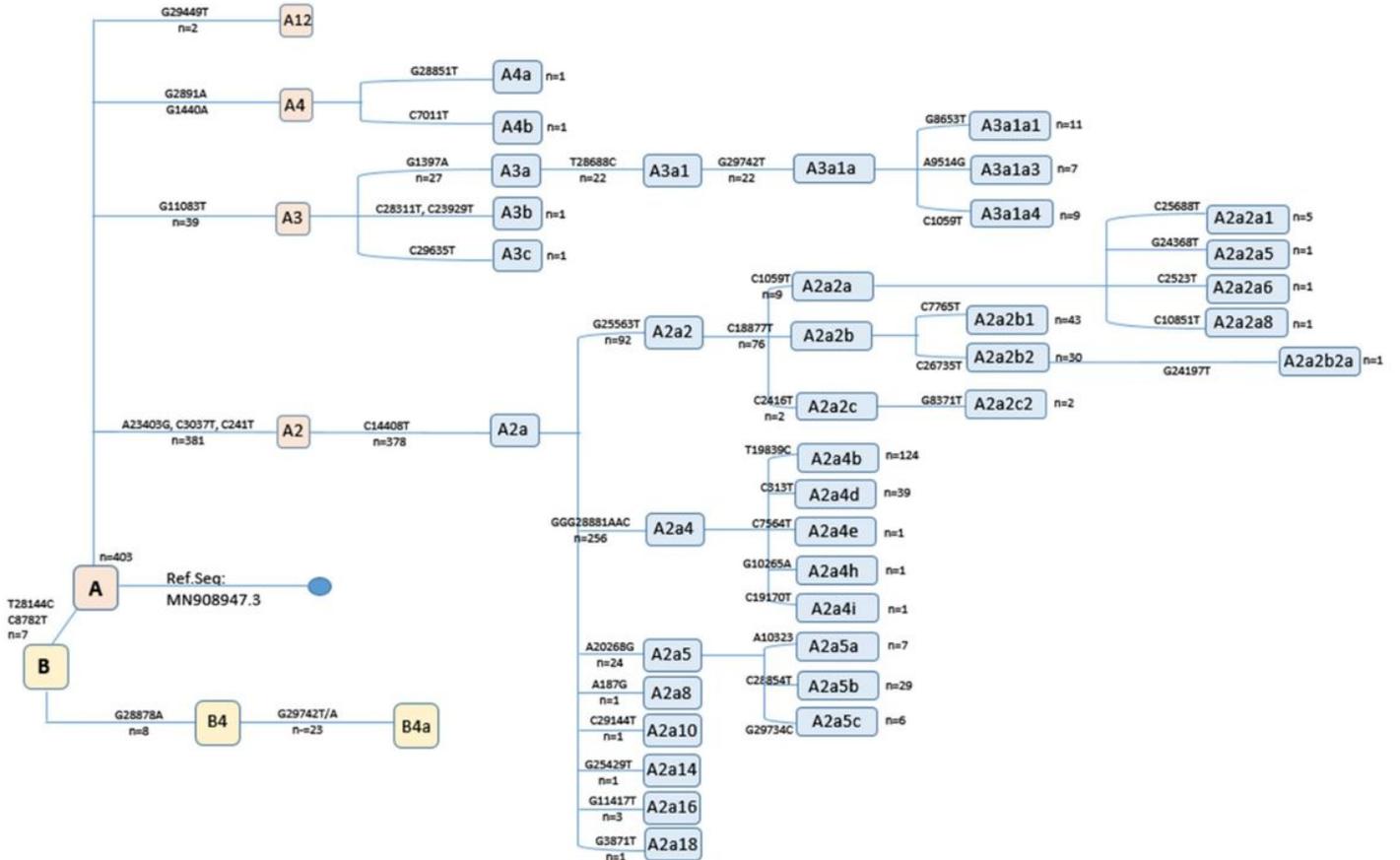


Figure 3

Distribution of the SARS-CoV-2 haplogroups in the study group. Branch-delineating variations and genome frequencies are indicated.

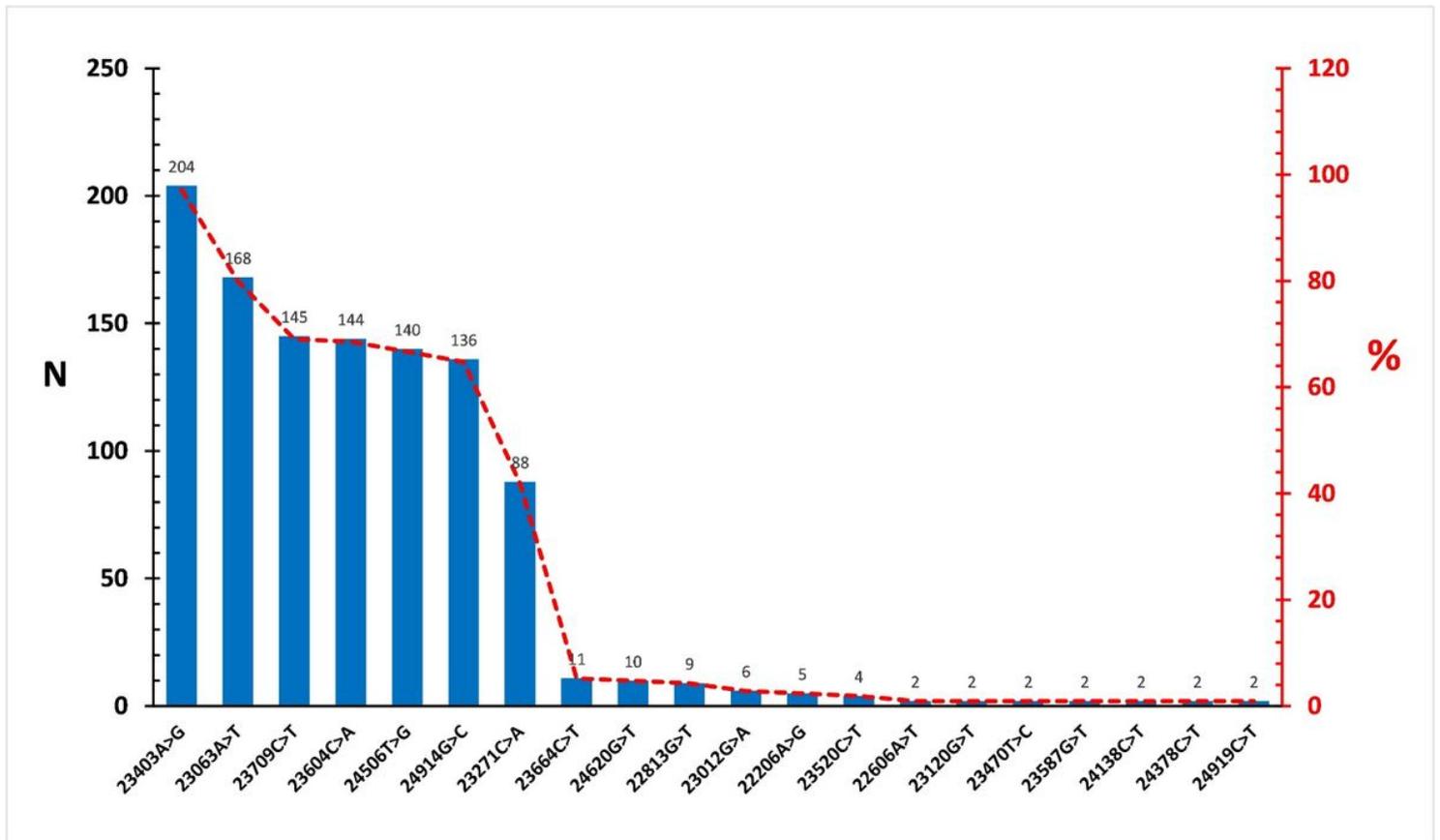


Figure 4

Distribution of the frequently-observed variations in spike region dataset.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement1.xls](#)