Mutational insights among the structural proteins of SARS-CoV-2; comprehensive analyses in the six continents

Mohammad Abavisani
Mashhad University of Medical Sciences

Karim Rahimian
Tarbiat Modares University

Bahar Mahdavi
Royan Institute for Biotechnology, ACECR

Samaneh Tokhanbigli
Islamic Azad University, Tehran

Mahsa Mollapour Sisakht
Erasmus University Medical Center

Amin Farhadi
Payame Noor University

Mansoor Kodori Kodori
Bam University of Medical sciences

Mohammadamin Mahmanzar
Kish International Campus University of Tehran

Zahra Meshkat (zmeshkat@gmail.com)
Mashhad University of Medical Sciences

Research Article

Keywords: Severe acute respiratory syndrome coronavirus 2, SARS-CoV-2, Spike protein, Envelope protein, Membrane protein, Nucleocapsid protein, structural proteins, Mutation

Posted Date: July 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1826408/v1

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

**Background:** Mutations among the structural proteins of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) can lead to the emergence of new variants with different features in mortality and sensitivity toward drugs and vaccines. Here we aimed to investigate the mutations among structural proteins of SARS-CoV-2 globally.

**Methods:** We analyzed samples of amino-acid sequences (AASs) for envelope (E), membrane (M), nucleocapsid (N), and spike (S) proteins from declaration the coronavirus 2019 (COVID-19) as pandemic to January 2022. Then, the existence of mutations and their locations have been considered by the sequence alignment to the reference sequence, categorized based on frequency and continent. Finally, the related human genes with the viral structural genes have been discovered, and their interactions have been reported.

**Results:** The results indicated that the most relative mutations among the E, M, N, and S AASs happened in the regions of 7 to 14, 66 to 88, 164 to 205, and 508 to 635 AAs, respectively. The most frequent mutations in E, M, N, and S proteins were concluded as T9I, I82T, R203M/R203K, and D614G. D614G is the most frequent mutation in all six geographical areas. Following D614G, L18F, A222V, E484K, and N501Y rank second to fifth most frequent mutations in S protein globally. Besides, A-kinase Anchoring Protein 8 Like (AKAP8L) has been shown as the linkage unit between M, E, and E cluster genes.

**Conclusion:** Screening the mutations of structural proteins can help scientists introduce better drug and vaccine development strategies.

Background

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a single-stranded, positive-sense RNA virus with a genome size of 29,903 nucleotides (1). The infectivity and mortality rate of coronavirus 2019 (COVID-19), the infectious disease caused by SARS-CoV-2 and has been declared as a pandemic on 11th March 2020, are variable among different countries and the exact factor influencing this variability is not lucid yet (2).

Despite the proofreading mechanisms of the virus, mutation rates of coronaviruses are between $10^{-5}$ and $10^{-3}$ substitutions per nucleotide site per cell infection (s/n/c); therefore, several mutations have been detected by wide-ranging sequencing (3, 4). The genome of SARS-CoV-2 contains 12 open reading frames (ORF) that encode 22 nonstructural proteins and 4 structural proteins including envelope (E) protein, membrane (M) protein, nucleocapsid (N) protein and spike (S) protein (5). The S protein, which plays key role in recognition by human host cell surface receptor angiotensin-converting enzyme 2 (ACE-2), consists of an N-terminal S1 subunit and a C-terminal S2 subunit. The occurrence of a point mutation in the S1 subunit named D614G mutation resulted in the dominant variant of SARS-CoV-2. This mutation is associated with the higher viral load, increased fitness and enhanced infectivity of the virus and leads to the emergence of disease in the younger age (3, 6-8). In fact, emergent mutations may
appear a new variant of concern (VOC) which achieves the increased potential in binding affinity to human receptors and higher infectivity and mortality power than previous variants. The B.1.617.2 (Delta) variant is an example of such demonstration which owns higher transmissibility and resistant to neutralization due to the combination of mutations occurred in the S protein (9). The B.1.1.529 (Omicron) variant is the fifth and up to now, the last VOC possessing about 40 mutations containing one mutation in E protein, 3 mutations in M protein, 6 mutations in N protein, and more than 30 changes in S protein. It has been identified in November 2021 in South Africa and increased infectivity and higher potential of stimulating the immune response was displayed in such variant by the virtue of high quantity of mutations (10). These findings highlight the importance of mutations pattern affecting the COVID 19 prevalence, infectivity, and mortality rate among different countries which needs to be elucidated.

At present, vaccination is the most advantageous strategy for combating COVID-19. Most of the designed vaccines inhibit viral pathogenesis via targeting the S protein (11, 12). Hence, investigating the effect of the mutations has an important role in determining the quality of vaccinations. On the other hand, some of the therapeutic strategies are based on the interaction between drugs and S proteins (13-15). Additionally, other structural proteins are considered to be attractive targets for manufactured drugs and vaccine formulations (16, 17). Consequently, identifying the geographical distributions and evolutionary trends of the structural proteins of SARS-CoV-2 triggers the intelligent approach in the epidemiological researches and molecular designing of the vaccines and drugs.

In the current study, we aimed to discover the frequencies of mutations among structural proteins of SARS-CoV-2 globally and also within the different continents, separately. Then, we addressed the evolutionary trends of the mutation's bedside considering their characteristics from the beginning of the pandemic to January 2022 and ultimately, related human genes with structural proteins of SARS-CoV-2 and the interactions were studied.

**Materials And Methods**

**Sequence retrieval**

The current study performed by evaluation of whole data belonging to 4 structural genes of SARS-CoV-2 AASs. All AASs were considered compared to the reference sequence. For the reference sequence, we utilized AAS of the Wuhan-2019 virus with access number ‘EPI_ISL_402124’. AASs of E region (with 75 AA owned by 6524654 samples), M region (with 222 AA owned by 6524705 samples), N region (with 419 AA owned by 6523549 samples) and S region (with 1273 AA owned by 6603870 samples) from January 2020 until January 2022 were extracted using GISAID database (https://www.gisaid.org). Access to this database was according to the permission of Erasmus Medical Center (https://www.gisaid.org/collaborations/collaboration-on-h5-antigenic-cartography) (18-20). The genomic position of evaluated amino acids comprises 26245 to 26472 for E protein, 26523 to 27191 for M protein, 28274 to 29533 for N protein and 21563 to 25384 for S protein (Fig. 1).

**SARS-CoV-2 sequences analyses**
Python 3.8.0 software was operated to initialize the FASTA files, extract the 4 mentioned structural proteins from other genes and conduct the SARS-CoV-2 sequences alignment and analysis in order to detecting the presence of any mutations. Each difference between sample and reference sequence was interpreted as a mutation and the location and substituted AA were reported. For each of the structural proteins, non-human samples, samples with different number AAs compared to reference and samples with non-specified AAs were excluded. The total process was optimized by applying 'Numpy' and 'Pandas' libraries.

The algorithm utilized for detecting mutants is as follows:

For the reason that all sequences have equal lengths, the following algorithm used ‘Refseq,’ and ‘seq’ refer to reference sequence and sample sequence, respectively.

```
For refitem, seqitem in zip (refseq, seq)
If (refitem != seqitem)
    Report a new mutant
```

After extraction of structural proteins, the continent name and geographical coordinates for each sample were achieved and reported using pycountry-convert 0.5.8 software and 'Titlecase' library in Python to depict the global prevalence maps of mutations. Lastly, global maps were drawn using Matlab 2021 'Geobubble' package.

**Data normalization and statistical analysis**

Data normalization was performed by usage of R 4.0.3 and Microsoft Power BI. The normalized frequency of each of studied regions was achieved due to the better comparison of the data attributed to each continent. Thus, the number of mutations was divided by the number of sequences on that continent that were comparable in equal proportions.

**Identification of related human genes and interactions**

To discover related human genes with E, M, N, S SARS-CoV-2 genes, The Human Protein Atlas database (https://www.proteinatlas.org/humanproteome/sars-cov-2) has been utilized to determine target genes. Additionally, STRING ver.11.5 with an average local clustering coefficient of 0.527 was utilized to determine the interaction between target genes. The interaction network by .tsv format was downloaded. Adjacent matrix was designed and imported to Cytoscape version 3.8 to visualize PPI network. Moreover, cytoHubba package was used to perform node ranking analysis to identify the hub genes.

**Results**

**Mutations quantity among geographical areas**
At the start, we decided to find the occurred mutations in order to understanding the incidence rates of mutations and realizing the potentially essential mutations statistically. The number of 6394483, 6177403, 5841477 and 895738 sequences belonging to E, M, N and S proteins were qualified and studied for identifying the number of mutations, respectively.

According to the achieved data, 96.40% of E amino-acid (AA) sequences (AASs) displayed no mutation (Fig 2). These features for M, N and S AASs were 36.76%, 2.20% and 2.11%, respectively. Additionally, the one mutation incidence rate was 3.56%, 59.64%, 5.68% and 26.86%, sequentially. 0.02%, 2.80%, 7.11% and 26.15% for E, M, N and S AASs displayed two mutations, respectively. Besides these data, the incidence rate of three and four mutations for E, M, N and S AASs has been demonstrated in Fig 2.

The achieved data belonging to the E protein demonstrated that 77.72% of AASs in Africa and 95.72% of Asia AASs did not display any mutation (Fig 2). This exhibition in Europe, North America, South America and Oceania has been observed in 96.10%, 97.44%, 97.30% and 96.86% of AASs, respectively. In comparison, 49% of Africa AASs did not include any mutation for M protein. Such feature for M protein was concluded by 36.44%, 36.62%, 35.03%, 60.45% and 39.04% in Asia, Europe, North America, South America and Oceania, successively. One mutation of M protein was shown in 47.35%, 61.19%, 60.27%, 60.36%, 37.87% and 56% in Asia, Europe, North America, South America and Oceania AASSs, respectively. 1.43%, 1.9% and 2.13% of Africa, Asia and Europe AASSs illustrated two mutations in M AASSs, same as the 4.01%, 1.33% and 2.35% of AASSs in North America, South America and Oceania, successively.

4.53% of the samples in Africa were without any mutation in N protein. Also 2.98%, 1.80%, 2.36%, 1.14% and 4.31% of N AASSs in the area of Asia, Europe, North America, South America and Oceania were without mutation occurrence, respectively. In contrast to Africa, which displayed 19.35% with one mutation, the one mutation incidence rate in other 5 areas was noticeably lower and except Oceania, other areas displayed almost similar one mutation incidence rate. The percentage of N AASSs with two mutations in Oceania and Africa were higher than other areas. This demonstration in Oceania and Africa AASSs were 28.81% and 16.55%, respectively but in Asia, Europe, North America and South America were 7.07%, 4.8%, 9.53% and 4.95%, successively. Concerning the S protein, it has been resulted that in South America, only 0.46% of AASSs did not display any mutation and about 82% demonstrated four and more mutations. For such protein, the no mutation incidence rate in Oceania was 8.45%, the highest no mutation incidence rate. The one mutation incidence rate among S AASSs has been demonstrated as 36.01%, 35.81%, 20.75%, 31.70%, 6.95% and 13.78% in the ASSs of Africa, Asia, Europe, North America, South America and Oceania, respectively. 81.96%, 35.92%, 24.07%, 17.14%, 16.99% and 2.52% of S AASSs displayed four and more mutations among South America, North America, Asia, Africa, Europe and Oceania ASSs, in order from large to small. The prevalence of AASSs with one mutation in Africa, Asia and Europe were higher than other types of achieved data (Fig 2). Besides, the most prevalent number of mutations in Oceania and Americas were two mutations and more than three mutations, respectively.

In the following, we decided to draw a heat map for mutations to detect their frequency in total and among each of the areas. Data displayed the most mutations relative to the total AASSs among the E, M, N
and S AASs occurred in the regions of 7 to 14 AA (0.0018 frequency), 66 to 88 AA (0.0279 frequency), 164 to 205 AA (0.0294 frequency) and 508 to 635 AA (0.0079 frequency), respectively (Fig 3). The second highest mutations frequency in the E, M, N and S AASs arose in the regions of 56 to 63 AA (0.0006 frequency), 1 to 22 AA (0.0010 frequency), 205 to 246 AA (0.0201 frequency) and 1 to 127 AA (0.0048 frequency), respectively. The necessity of heat map refers to the variation in the appearance number of mutations and dispersion in samples of which mutations have been occurred.

**Mutation's features based on the geographical areas**

In the next step, the locations of mutations in the protein structure and their frequency have been considered to identify more dimensions of mutations. As it has been mentioned in the Fig 4, the most frequent mutation in the E protein is attributed to T9I with the frequency rate of 0.0128 and after that, P71L with 0.0068 frequency, V62F 0.0066 frequency, L21F/L21V with 0.0017/0.0003 frequencies and V58F with 0.0013 frequency have the highest frequency rate of mutations. Locations of top three frequent mutations were shown in Fig 5 section A. Accordingly, T9I were the most frequent mutation in Europe, Oceania, North America and South America with the 0.0187, 0.0249, 0.0066 and 0.0049 frequency rates, respectively. Nevertheless, P71L was the most frequent mutation in Africa and Asia with frequency rates of 0.1643 and 0.0146, respectively. V62F is one of the first ten frequent mutations in Asia (0.0118 frequency), Europe (0.0016 frequency) and North America (0.0024 frequency), in contrast to Africa (0.0011 frequency), Oceania (0.0004 frequency) and South America (0.0012 frequency) which this mutation was as eighth, sixth and ninth, respectively.

Regarding the M protein, analysis showed I82T (0.6015 frequency), D3G (0.0077 frequency), A63T (0.0073 frequency), Q19E (0.0072 frequency) and A2S (0.0033 frequency) are the first five mutations with highest frequency, respectively (Fig 4). I82T was the most frequent mutation in all six areas (Fig 5 section B). This situation is different from D3G mutation which is not the second frequent mutation in Asia and North America. In these areas, F28L (0.020 frequency) and A81S (0.0083 frequency) were at the second position of frequent mutations, respectively. In Africa and Europe, A63T was the third frequent mutation with 0.0224 and 0.0091 frequencies, respectively. On the other hand, the third frequent mutation in Asia, Oceania, North America and South America were D3G (0.0048 frequency), Q19E (0.0259 frequency), S197N (0.0069 frequency) and R164H (0.004 frequency), respectively.

Analysis of N AASs data illustrated that the R203M/R203K with 0.6084/0.2489 frequencies was at the first position of frequent mutations (Fig 5 section C). Globally, D377Y (0.6134 frequency) mutation ranks second, D63G (0.6002 frequency) ranks third, G215C (0.5479 frequency) ranks fourth and G204R/G204P (0.2352/0.0134 frequencies) ranks fifth mutation (Fig 4). In all continents except South America, up to fourth position of frequent mutations were similar to the global results. Analysis data of South America resulted in the different arrangement. The frequency of R203M mutation is higher than R203K mutation in all continents excluding South America. The R203M/R203K frequencies in Africa were 0.4195/0.1965, in Asia were0.6033/0.3052, in Europe were0.6074/0.2826, in North America were 0.6310/0.1776 and in Oceania were 0.6143/0.3008. However, in South America the frequencies of R203M/R203K were
A further dimension of differences between South America and other areas belongs to the positions of second and third frequent mutations. G204R (0.5685 frequency) and P80R (0.4184 frequency) ranks second and third mutations in South America.

The pattern of mutation frequency for S AASs displayed that in the world, D614G with 0.9756 frequency achieved first place among frequent mutations. In the following, L18F (0.1680 frequency), A222V (0.1579 frequency), E484K (0.1454 frequency) and N501Y (0.1120 frequency) ranks second to fifth frequent mutations (Fig 4). The first frequent mutation in S AASs is identical in all six geographical areas; however, the frequencies are different between them (Fig 5 section D). The frequency of D614G in Africa, Asia, Europe, North America, South America and Oceania was resulted as 0.8884, 0.9579, 0.9743, 0.9835, 0.9959 and 0.9047, respectively. In Africa, P681R/P681H mutations (0.0674/0.0396 frequencies) ranks second, Q677H/Q677K mutations (0.0661/0.0163 frequencies) ranks third, L452R mutation (0.0728 frequency) ranks fourth and S477N mutation (0.0712) ranks fifth. In addition to the point that the second mutation in Asia attributed to E484K/E484Q (0.1112/0.0211 frequencies), P681R/P681H (0.1127/0.0150 frequencies), W152L (0.0897 frequency) and G769V (0.0903 frequency) are the third to fifth mutations in S AASs. In Europe, A222V (0.4217 frequency) is the second frequent mutation and L18F (0.1702 frequency), S477N (0.0999 frequency) and S98F (0.0466 frequency) are the third to fifth, respectively. Among the mutations occurred in S AASs from North America, E484K (0.1393 frequency) ranks second, L18F (0.1227 frequency) ranks third, P681H/P681R (0.0849/0.0268 frequencies) ranks fourth and L452R (0.1021) ranks fifth which are different in the types in frequent mutations from the South America. In South America, V1176F with 0.8592 frequency, E484K with 0.8268 frequency, N501Y with 0.7745 frequency and L18F with 0.7742 frequency are the frequent mutations, respectively. S477N (0.6864 frequency), P681R (0.0229 frequency), V1068F (0.0196 frequency) and N439K (0.0179 frequency) are the frequent mutation following D614G mutation in Oceania, respectively. Additional data have been listed in Additional file 1, 2, 3 and 4.

Evolutionary trends of emergence and distribution of top ten mutations with respect to the time and geographical areas

In order to more practical approach to mutations, identifying the trend of mutations emergence and the timeline of their spreading can be considered. The current part of study can be efficient in recognize the factors influencing the efficiency or inefficiency of drugs and vaccines. The distribution pattern of each of the top ten mutations, based on the time of collection, has been illustrated in Fig 6 and supplemental data have been shown in Additional file 5, 6, 7 and 8.

The T9 mutation, which is the most frequent mutation in E AASs in the world, began to prevail from October 2021 and till January 2022, it is present with 0.0693 frequency rate. In comparison, P71 mutation gained in prevalence in May 2020 and after decreasing in August 2020, restarted to increase from September 2020. P71 mutation was at maximum frequency rate in March 2021 with 0.0257 frequency rate. Although V62 mutation was present from the first days of pandemic, it has been increased from August 2021 and in October 2021 was at maximum frequency rate (0.0058). In Africa, the emergence of
P71 has been detected in August 2020. In subsequent, it started to increase up to April 2021 with highest frequency rate (0.3673) and gained to decrease till September 2021. Accordingly, the highest frequency of P71 mutation in all other continents is almost identical to Africa. In Asia, V62 mutation increased noticeably from August 2021 and displayed its maximum frequency rate (0.0901) in October 2021. At the beginning of pandemic in South America, V58 mutation has been grown and was at highest frequency rate (0.0457) in May 2020. However, it declined from July 2020.

The most worldwide frequent mutation of M AASs, I82, has notably frequency rate in January 2021 (0.1095). The second globally peak of I82 prevalence started from May 2021 and it has the highest frequency rate (0.9969) in October 2021 (Fig 6). From this perspective, except South America, the evolutionary trend in distribution of I82 mutation in all continents have almost identical pattern. Although I82 was detectable in South America from the beginning of pandemic, it has almost consistent and near zero frequency rate before April 2021 and started to prevail from July 2021. In all areas, the Q19 mutation increased from October 2021; at the time which I82 mutation began to decrease in the entire world.

The prevalence of mutations among N AASs is fluctuant. R203 mutation has a peak of frequency rate in January 2020 and began to increase from February 2020 till August 2020. It globally started to prevail again from November 2020 and has a growing pattern in the following with 0.9907 frequency rate in January 2022 (Fig 6). Considering the evolutionary trends in areas resulted in the similar pattern of R203 prevalence excluding South America. The achieved data from this area demonstrated almost steady pattern of frequency rate for R203 from April 2020 to January 2022 the evolutionary trends of D63 and G215 mutations have approximately identical pattern and both started to increase from April 2021 in all continents. Also, one of the frequent and exclusive mutations in South America, P80 mutation, has been increased from November 2020 and started to decrease from June 2021.

The growing evolutionary trend of D614 mutation has been started from February 2020 in the entire world. In contrast to other areas, in Africa the mentioned mutation did not have steady pattern and has been demonstrated fluctuant. L18 mutation has been increased from August 2020 and began to decrease from July 2021 in the world. Such pattern has been demonstrated by E488 and N501, globally (Fig 6). Contrary to them, A222 mutation displayed different trend. It gains in prevalence from July 2020 and started the decreasing pattern from October 2020. In May 2020, S477 mutation, which was one of the top ten frequent mutations in Oceania, began to increase and decreased just three months later (August 2020). The results achieved from South America demonstrated that except D614, approximately all other mutations obviously started to increase from November 2020.

**Protein-Protein Interaction (PPI) network presentation**

The Protein-Protein Interaction (PPI) network with 57 nodes and 153 edges presents the interaction between E, M, N, S SARS-CoV-2 protein and human proteins (Fig 7) (See Additional file 9). Through the ranking analysis, Ras GTPase-activating protein-binding protein 1 (G3BP1) was identified as high human gene rank (Fig 8). Additional data have been illustrated in Additional file 10. The network showed the linkage between the M protein cluster gene and E and N members which are linked with the A-kinase
anchoring protein 8 like (AKAP8L) human gene playing role as a bottleneck. Also in this achieved network, Zinc Finger DHHC-Type Palmitoyltransferase 5 (ZDHHC5) and Golgin A7 (GOLGA7) have been shown as the human genes with highest interaction with S protein.

Discussion

From the beginning days of March 2020 and announcement of the COVID-19 as a pandemic, it has been prevailed globally with variable demonstrations among countries. This variability is due to the differences in infectivity and mortality potentials of prevailed virus influenced by distinct factors including community characteristics such as age, genetic basis and mutations among predominant virus (21). Vaccination for prevention and therapeutic strategies for treatment are two important tools to inhibiting the global disease. Structural proteins have been persuading scientists to be targets for drug and vaccine developments (22, 23). Therefore, considering the mutations in these proteins and identifying their impacts on functions will be efficient in high-quality production and development of such prevention and remedial tools.

To focus on the S protein as a pathogenicity-initiator and antibody-stimulator influenced identifying of new mutations emergence through S protein and considering the potential of drug and vaccine efficacies for new variants. N501Y was one of the 17 mutations in the viral genome attributed to B.1.1.7 (Alpha) variant which has enhancing effect in viral attachment and infectivity (24). N501Y appearance was reported in UK in December 2020 and entered the USA in at the end of December 2020. The evolutionary trend belonging to the mutation emergence of S AA position 501 is consistent with these time periods. The mentioned mutation is associated significantly with the enhancement of binding affinity and consequently, increasing of viral infectivity (25). Mutation at the S AA position 501 besides D614G and E484K were two of nine mutations detected in B.1.351 (Beta) variant which reported in the second wave of COVID-19 in South Africa in October 2020. Accumulation of such mutations, which concluded as the members of top ten frequent mutations globally in our study, not only elicited the increased potential of transmission, also was efficient in reducing the neutralization by monoclonal antibody therapy (26). The B.1.1.28 (Gamma) variant, which detected for the first time in Brazil in December 2020, harbors ten mutations in the S protein including L18F, T20N, P26S, D138Y, R190S, H655Y, T1027I, V1176, K417T, E484K, and N501Y (27). According to the current study, all of these mutations were detected as the top ten frequent mutations in the continent of South America excluding the mutation at the S AA position 138. Mutations at the positions of 18, 417 and 484 in the receptor-binding domain (RBD) are the main factors in enhancing the viral binding in both Beta and Gamma variants. The next VOC, Delta variant, was the main reason of the COVID-19 wave in April 2021 among Indian society. In addition to T19R, T478K, P681R and D950N, Delta variant owns six other mutations and similar to Gamma variant, it harbors ten mutations located in S protein in total (28). The entire first four mentioned mutations were demonstrated in the current study as the members of frequent mutations in Asia which has been increased noticeably after April 2021. T19R, T478K, P681R and D950N mutations help the Delta variant in escaping immune response system, increasing the viral attachment and enhancing viral replication, respectively (29, 30). In addition to Delta variant, mutation at the AA position 681 belonging to the S protein has been detected in
the prior variants as well. P681H is one of the observed mutations in Alpha variant. According to our results, the change from proline to histidine is the most frequent mutation attributed to the S AA position 681 in North America and concluded as the second frequent mutation at the S AA position 681 in other continents except Europe and South America. P681R is more frequent than P681H in Asia, Africa and Oceania. Similarly, because of the role of both mutations in increasing the cleavage of S1/S2 site within the S protein by furin enzyme, these mutations have positive correlation with increased viral infectivity and virulence of COVID-19 (31, 32) . L452R are another mutation of Delta variant in the RBD of S protein. In the current study, this mutation has been shown as one of the top ten frequent mutations in Asia as well as Africa, North America and Oceania. Co-occurrence of L452R beside T478K can enhance the viral transmissibility via influence on viral attachment. Summation of such mutations is related to Delta variant. Similarly, in Beta and Gamma variants, E484K/N501Y double mutations in RBD have same outcomes (33) . The number of mutations occurring in Omicron variant, which is the last VOC so far, is increasing daily and because of such mutations especially in S protein, it is believed that the restricting the pandemic situation and end of fatal demonstrations of COVID-19 is nearer than previous (10, 34) . Our data analyzing showed ongoing frequencies of two mutations in N protein related to Alpha, Gamma and Omicron variants, G204R and R203K, in approximately near the time of variant identification in Africa. Furthermore, it has been displayed that the frequency of R203M mutation is higher than R203K mutation in all continents but South America during the period of research. Co-occurrence of these two mutations has positive correlation with the potential of infectivity and viral virulence through increasing viral replication (35) . On the other hand, D614G was displayed as the most frequent mutation in S protein in the world and independently, in each of the continents. Changing the aspartic acid to glycine resulted to enhancing the binding affinity to ACE-2. This mutation in S protein has positive effect on the transmissibility increase and is related to the greater viral load (3) . In spite of these outputs, there is not any evidence about D614G and increased severity of COVID-19 (3, 36) . Even though the sensitivity to antibody neutralization become increased by the virtue of the mutation, higher infectivity and transmissibility potential besides lack of effect on the strength of severity are the cause of being the most frequent mutation in S protein globally.

M protein is another structural protein that has not been attracted the scientists as much as it should. Despite it is high conserved between SARS-CoV and SARS-CoV-2, the role in viral assembly and viral pathogenicity is important enough to investigate the mutations (37) . I82T mutation has been observed as the most frequent mutation globally. The similar results have been concluded about mutations frequencies in another study researching the M mutations in a geographical region (38) . Additionally, Q19E mutation of M protein in Omicron variant displayed increasing trend of frequency in November 2021 mainly in Africa and globally.

E protein entitled as the smallest structural protein of SARS-CoV-2 has critical role in viral assembly and also pathogenesis. Previous studies concluded the same we observed in the point that evolutionary trend of E mutations were slow-going (39-41) . T9I mutation, which concluded as the most frequent mutation globally in our study, has the increased frequency in about November 2021. This mutation is one of the mutations occurred in Omicron variant. The global pattern of mutations emergence and distribution
concluded in our study acknowledges slower progress of E mutations compared to others. According to our results, 96.40% of E AASs did not possess any mutation. Due to the high conservancy attributed to this structural protein, it might be attractive for the vaccine and drug researchers as a potential target in COVID-19 (42).

Although S protein attracted the main attentions, to consider the other structural proteins can have the similar to or may be higher results in drug and vaccination development than the S protein. Recently, new serological method for detection of SARS-CoV-2 has been introduced using the N protein (34, 43). Because the conserved regions in N protein of SARS-CoV-2 Harbor the highest AA similarity with SARS-CoV, researchers should investigate more specific regions if they are curious about determining new methods based on the SARS-CoV-2 N protein (44). Also, mutations can impact on the results of diagnostic tests. T135I mutation can result to false negative using rapid antigen test. Although this mutation was not classified among top ten frequent mutations of N protein in the current study, it is urgent to screen the mutations accurately to product and develops the reliable diagnostic tests. Therapeutic strategies targeting the structural proteins must be according to the mutations and frequencies and functions due to them. Some mutations among S protein such as E484K and L452R resulted in the reduced sensitivity toward antibodies and requiring higher titers of convalescent sera for effective neutralization (9, 45, 46). N501Y mutation is also related to decreasing sensitivity against some specific antibodies such as COVA1-12 and CB6 in spite of the fact that it has not any pronounced effect on the neutralizing activity of plasma or sera from vaccinated persons (47, 48).

Identifying the related human genes and proteins with SARS-CoV-2 genes can enhance the scope of our finding about the host targets of such virus. According to the previous proteomic studies, ZDHHC5 and GOLGA7 were two human genes with high interaction with the S protein of SARS-CoV-2 (49). ZDHHC5 is localized in the endoplasmic reticulum (ER) or Golgi apparatus and has an important role in catalyzing protein palmitoylation. Also, GOLGA7 is another one of the cell palmitoylation system which plays role in regulating the ZDHHC enzyme activity (50, 51). Thus, ZDHHCs enzyme and GOLGA7 are essential for palmitoylation of viral proteins and with the last finding about their high confidence interaction with S protein of SARS-CoV-2, they can be introduced as a potential drug target. In confirmation of these findings, high interaction of ZDHHC5 and GOLGA7 with S protein has been concluded in the current study. Furthermore, among our analyses, G3BP1 has been introduced as the human gene with highest rank among all human genes. It has been reported that N protein of SARS-CoV-2 can interact with this protein in order to inhibiting the host stress granule (SG) assembly and promoting the viral infection (52). SGs are dynamic ribonucleoprotein (RNP) assemblies constructed in the stressful conditioned such as oxidative stress and viral infection. G3BP1 with G3BP2 are key SG-nucleating factors which have been selected as targets of viral proteins by interaction with their N-terminal nuclear transport factor 2-like (53, 54). The PPI network also displayed AKAP8L as a linkage unit between M protein cluster and E and N members. AKAP8 is associated with the family of structurally diverse scaffold proteins and AKAP8L is the homologue of that protein. They have been presented as the proteins influencing the tumorigenesis and tumor metastasis of cancer cells (55, 56). Although PPI network resulted in the report of related human genes with structural genes of SARS-CoV-2, the accurate effects of ZDHHC5 and GOLGA7 on S
protein of SARS-CoV-2, the mechanism of interaction between N protein and SG disassembly and also the exact role of AKAP8L in viral infection and its mechanism are unclear and more molecular research must be performed to explain the molecular processes.

Our study encountered two limitations. First, in this research we considered AASs without studying the nucleotide sequences. This leads to the not checking the other aspects of new-emerging variants such as codon bias. Second limitation attributed to variation in samples and mutations reporting between different regions. Differences in mutation reports leads to the ignoring the present but not discovered mutations and consequently, it may cause not studying the new variants.

Conclusion

Considering the frequencies of functional mutations and the rate of their occurrence brought about the fact that the rate of vaccination and drug development against COVID-19 is subsequent to the virus mutations. Therefore, the faster the mutations are screened, the faster development strategies are implemented to effectively combat COVID-19.

Abbreviations

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; AASs: Amino-acid sequences; E: Envelope; M: Membrane; N: Nucleocapsid; S: spike; COVID-19: coronavirus 2019; AKAP8L: A-kinase Anchoring Protein 8 Like; S/N/C: Substitutions per nucleotide site per cell infection; ORF: Open reading frames; ACE-2: Angiotensin-converting enzyme 2; VOC: Variant of concern; PPI: Protein-Protein Interaction; G3BP1: GTPase-activating protein-binding protein 1; ZDHHC5: Zinc Finger DHHC-Type Palmitoyltransferase 5; RBD: Receptor-binding domain; ER: Endoplasmic reticulum; SG: Stress granule; RNP: Ribonucleoprotein; GISAID: Global Initiative on Sharing All Inuenza Data.

Declarations

Acknowledgments

The authors thank all of the researchers who have shared genome data openly via the Global Initiative on Sharing All Inuenza Data (GISAID).

Author contributions

M.M, K.R, M.M.S, S.T, A.F contributed to data collection, K.R, M.A, M.M, M.M.S contribute to study design, K.R, M.M design workflow and code, data analysis. K.R, M.M, B.M data visualization. M.A, S.T, M.K wrote the manuscript. M.M, K.R, monitored the accuracy of Additional data. B.M designed graphical contents. M.M and Z.M have edited and supervised the work. All authors reviewed the manuscript.
Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Availability of data and materials

All our study’s steps were performed per the Declaration of Helsinki (57). We analyzed the SARS-CoV-2 AASs provided in the GISAID database (https://gisaid.org) with the permission and monitoring of Erasmus Medical Center. The datasets supporting the conclusions of this article are included within the article and its supplementary files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Badua C, Baldo KAT, Medina PMB. Genomic and proteomic mutation landscapes of SARS-CoV-2. Journal of medical virology. 2021;93(3):1702-21.
2. World Health Organization. https://www.who.int/news-room/detail/27-04-2020-who-timeline—covid-19. 2020.
3. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020;182(4):812-27. e19.
4. Dolan ME, Hill DP, Mukherjee G, McAndrews MS, Chesler EJ, Blake JA. Investigation of COVID-19 comorbidities reveals genes and pathways coincident with the SARS-CoV-2 viral disease. Scientific Reports. 2020;10(1):20848.
5. Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerging microbes & infections. 2020;9(1):221-36.
6. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell. 2020;182(5):1284-94. e9.
7. Weissman D, Alameh M-G, de Silva T, Collini P, Hornsby H, Brown R, et al. D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. Cell host & microbe. 2021;29(1):23-31. e4.
8. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O’Toole Á, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 2021;184(1):64-75.e11.

9. Laurini E, Marson D, Aulic S, Fermeglia A, Pricl S. Molecular rationale for SARS-CoV-2 spike circulating mutations able to escape bamlanivimab and etesevimab monoclonal antibodies. Scientific Reports. 2021;11(1):20274.

10. Gupta S, Saarikko M, Pfützner A, Räisänen IT, Sorsa T. Compromised periodontal status could increase mortality for patients with COVID-19. Lancet Infect Dis. 2022;22(3):314.

11. Berguer PM, Blaustein M, Bredeston LM, Craig PO, D’Alessio C, Elias F, et al. Covalent coupling of Spike’s receptor binding domain to a multimeric carrier produces a high immune response against SARS-CoV-2. Scientific Reports. 2022;12(1):692.

12. Shi R, Shan C, Duan X, Chen Z, Liu P, Song J, et al. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. Nature. 2020;584(7819):120-4.

13. Chan WKB, Olson KM, Wotring JW, Sexton JZ, Carlson HA, Traynor JR. In silico analysis of SARS-CoV-2 proteins as targets for clinically available drugs. Scientific Reports. 2022;12(1):5320.

14. Strobell R, Adler J, Paran N, Yahalom-Ronen Y, Melamed S, Politi B, et al. Imatinib inhibits SARS-CoV-2 infection by an off-target-mechanism. Scientific Reports. 2022;12(1):5758.

15. León G, Herrera M, Vargas M, Arguedas M, Sánchez A, Segura Á, et al. Development and characterization of two equine formulations towards SARS-CoV-2 proteins for the potential treatment of COVID-19. Scientific Reports. 2021;11(1):9825.

16. Peng Y, Du N, Lei Y, Dorje S, Qi J, Luo T, et al. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. The EMBO journal. 2020;39(20):e105938.

17. Oliveira SC, de Magalhães MTQ, Homan EJ. Immunoinformatic Analysis of SARS-CoV-2 Nucleocapsid Protein and Identification of COVID-19 Vaccine Targets. Frontiers in immunology. 2020;11:587615.

18. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID’s Role in Pandemic Response. China CDC Weekly. 2021;3(49):1049.

19. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. Global challenges. 2017;1(1):33-46.

20. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2017;22(13).

21. Zhou B, Kojima S, Kawamoto A, Fukushima M. COVID-19 pathogenesis, prognostic factors, and treatment strategy: Urgent recommendations. Journal of medical virology. 2021;93(5):2694-704.

22. Peng Y, Du N, Lei Y, Dorje S, Qi J, Luo T, et al. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. The EMBO journal. 2020;39(20):e105938.

23. Oliveira SC, de Magalhães MT, Homan EJ. Immunoinformatic analysis of SARS-CoV-2 nucleocapsid protein and identification of COVID-19 vaccine targets. Frontiers in immunology. 2020:2758.
24. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science (New York, NY). 2021;372(6538).
25. Kuzmina A, Khalaila Y, Voloshin O, Keren-Naus A, Boehm-Cohen L, Raviv Y, et al. SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. Cell host & microbe. 2021;29(4):522-8. e2.
26. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv : the preprint server for biology. 2021.
27. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DdS, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science (New York, NY). 2021;372(6544):815-21.
28. Tian D, Sun Y, Zhou J, Ye Q. The Global Epidemic of the SARS-CoV-2 Delta Variant, Key Spike Mutations and Immune Escape. Frontiers in immunology. 2021;12.
29. Allen H, Vusirikala A, Flannagan J, Twohig KA, Zaidi A, Chudasama D, et al. Household transmission of COVID-19 cases associated with SARS-CoV-2 delta variant (B.1.617.2): national case-control study. The Lancet Regional Health-Europe. 2022;12:100252.
30. Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira IATM, et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. Nature. 2021;599(7883):114-9.
31. Saito A, Irie T, Suzuki R, Maemura T, Nasser H, Uriu K, et al. SARS-CoV-2 spike P681R mutation, a hallmark of the Delta variant, enhances viral fusogenicity and pathogenicity. bioRxiv : the preprint server for biology. 2021.
32. Sohail A, Tunc S, Nutini A, Arif R. Furin and the adaptive mutation of SARS-COV2: a computational framework. Modeling Earth Systems and Environment. 2021.
33. Sinha S, Tam B, Wang SM. RBD Double Mutations of SARS-CoV-2 Strains Increase Transmissibility through Enhanced Interaction between RBD and ACE2 Receptor. Viruses. 2021;14(1).
34. Yu J, Qin Z, Liu X, He X, Yao J, Zhou X, et al. High-specificity targets in SARS-CoV-2 N protein for serological detection and distinction from SARS-CoV. Computers in biology and medicine. 2022;143:105272.
35. Wu H, Xing N, Meng K, Fu B, Xue W, Dong P, et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. Cell host & microbe. 2021;29(12):1788-801. e6.
36. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell. 2021;184(1):64-75. e11.
37. Lu S, Ye Q, Singh D, Cao Y, Diedrich JK, Yates JR, 3rd, et al. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. Nature communications. 2021;12(1):502.
38. Shen L, Bard JD, Triche TJ, Judkins AR, Biegel JA, Gai X. Emerging variants of concern in SARS-CoV-2 membrane protein: a highly conserved target with potential pathological and therapeutic implications. Emerg Microbes Infect. 2021;10(1):885-93.
39. Badua CLD, Baldo KAT, Medina PMB. Genomic and proteomic mutation landscapes of SARS-CoV-2. Journal of medical virology. 2021;93(3):1702-21.

40. Islam MR, Hoque MN, Rahman MS, Alam A, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. Scientific reports. 2020;10(1):1-9.

41. De Maio F, Cascio EL, Babini G, Sali M, Della Longa S, Tilocca B, et al. Improved binding of SARS-CoV-2 Envelope protein to tight junction-associated PALS1 could play a key role in COVID-19 pathogenesis. Microbes and infection. 2020;22(10):592-7.

42. Nguyen TT, Pathirana PN, Nguyen T, Nguyen QVH, Bhatti A, Nguyen DC, et al. Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus). Scientific Reports. 2021;11(1):3487.

43. Algaissi A, Alfaleh MA, Hala S, Abujamel TS, Alamri SS, Almabhoub SA, et al. SARS-CoV-2 S1 and N-based serological assays reveal rapid seroconversion and induction of specific antibody response in COVID-19 patients. Scientific Reports. 2020;10(1):16561.

44. Maldonado LL, Bertelli AM, Kamenetzky L. Molecular features similarities between SARS-CoV-2, SARS, MERS and key human genes could favour the viral infections and trigger collateral effects. Scientific Reports. 2021;11(1):4108.

45. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science (New York, NY). 2020;369(6506):1014-8.

46. McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, et al. SARS-CoV-2 immune evasion by the B. 1.427/B. 1.429 variant of concern. Science (New York, NY). 2021;373(6555):648-54.

47. Rees-Spear C, Muir L, Griffith SA, Heaney J, Aldon Y, Snitselaar JL, et al. The effect of spike mutations on SARS-CoV-2 neutralization. Cell reports. 2021;34(12):108890.

48. Zhu X, Mannar D, Srivastava SS, Berezuk AM, Demers J-P, Saville JW, et al. Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. PLoS biology. 2021;19(4):e3001237.

49. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459-68.

50. Gadalla MR, Veit M. Toward the identification of ZDHHC enzymes required for palmitoylation of viral protein as potential drug targets. Expert opinion on drug discovery. 2020;15(2):159-77.

51. Salaun C, Locatelli C, Zmuda F, Cabrera González J, Chamberlain LH. Accessory proteins of the ZDHHC family of S-acylation enzymes. Journal of Cell Science. 2020;133(22):jcs251819.

52. Luo L, Li Z, Zhao T, Ju X, Ma P, Jin B, et al. SARS-CoV-2 nucleocapsid protein phase separates with G3BPs to disassemble stress granules and facilitate viral production. Science Bulletin. 2021;66(12):1194-204.

53. White JP, Lloyd RE. Regulation of stress granules in virus systems. Trends in microbiology. 2012;20(4):175-83.
54. Guillén-Boixet J, Kopach A, Holehouse AS, Wittmann S, Jahnel M, Schlüßler R, et al. RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. Cell. 2020;181(2):346-61. e17.

55. Taskén K, Aandahl EM. Localized effects of cAMP mediated by distinct routes of protein kinase A. Physiological reviews. 2004;84(1):137-67.

56. Luo QY, Di T, Qiu MZ, Xia ZF, Du Y, Lin RD, et al. High AKAP8L expression predicts poor prognosis in esophageal squamous cell carcinoma. Cancer cell international. 2022;22(1):90.

57. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama. 2013;310(20):2191-4.

Figures

Figure 1

The genomic position of 4 structural proteins of SARS-CoV-2; E protein includes 26245 to 26472. Position of M, N and S proteins are 26523 to 27191, 28274 to 29533 and 21563 to 25384, respectively

Figure 2

Pie chart plot of the number of mutations among structural proteins of SARS-CoV-2 up to January 2022. The incidence rate of one, two, three, four and more mutations in addition to the rate of lack of any mutation among E, M, N and S proteins has been displayed in the clusters of A to D, respectively. 96.40% of E AASs, 36.76% of M AASs, 2.20% of N AASs and 2.11% of S AASs clustered as the AASs without any mutation entire the world. The rate of four and more mutations among the M, N and S AASs calculated as 0.01%, 76.54% and 30.62% globally. This index for E AASs did not display in the current pie chart due to the very low amount of attributed index ( = 3.75E-0.004% )

Figure 3

The heat map of mutations among structural proteins of SARS-CoV-2 as of January 2022; these indicate the rate of mutation per 100 amino acids. The highest frequency rate among the E, M, N and S AASs occurred in the regions of 7 to 14 AA, 66 to 88 AA, 164 to 205 AA and 508 to 635 AA, respectively. In the
world, the region of 56 to 63 for the E AASs has been concluded as the position with second relative frequent mutations. The position with same feature were 1 to 22, 205 to 246 and 1 to 127 among the M, N and S AASs globally

**Figure 4**

Top ten mutations with the highest frequency worldwide and geographic areas; the position of altered AAs and substituted ones are shown differently based on the frequency percentage of substituted AA. For better data representation, the data is represented by a logarithm based on 10. In total, T9I, P71L, V62F, L21F/L21V, V58F, L73F, S68F, S55F, V49L and A41V were displayed as the ten mutations with highest frequency rate of mutations for E AASs. I82T, D3G, A63T, Q19E, A2S, V70L, A81S, F28L, S197N and T30I are the top ten frequent mutations among M AASs globally. Additionally, R203M/R203K, D377Y, D63G, G215C, G204R/G204P, D3L, S235F, Q9L, A220V and P199L ranks first to tenth frequent mutations for M AASs entire world and also, top ten frequent mutations for S AASs were concluded as D614G, L18F, A222V, E484K, N501Y, V1176F, T1027I, D138Y, P26S and T20N globally
Figure 5

The locations of the top three frequent mutations occurred in E protein (A), M protein (B), N protein (C) and S protein (D) have been demonstrated. The mutations are among the top frequent mutations in the total
Figure 6

Timeline for report of mutations and evolutionary trends of top ten high-rate mutations in the structural proteins of SARS-CoV-2 among different geographic areas including North America, South America, Europe, Asia, Oceania, and Africa and globally.

Figure 7

The visualization of PPI network with 57 nodes and 153 edges.

Figure 8

Hub genes identification was concluded by node ranking analysis. As it is seen, the AKAP8L human gene is the linkage gene between the M protein cluster gene and E and N genes of SARS-CoV-2. Moreover, ZDHHC5 and GOLGA7 showed high interaction with the S protein of SARS-CoV-2.
Supplementary Files

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

- Additionalfile1.xlsx
- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile5.xlsx
- Additionalfile6.xlsx
- Additionalfile7.xlsx
- Additionalfile8.xlsx
- Additionalfile9.csv
- Additionalfile10.csv