Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Genetic variations among SARS-CoV-2 strains isolated in China

Eltayib H. Ahmed-Abakur\textsuperscript{a,b}, Tarig M.S. Alnour\textsuperscript{a,b,*}

\textsuperscript{a} University of Tabuk, Faculty of Applied Medical Sciences, Department of Medical Laboratory Sciences, Tabuk, Saudi Arabia
\textsuperscript{b} AlZaiem AlAhhari University, Faculty of Medical Laboratory Sciences, Department of Microbiology and Immunology, Khartoum North, Sudan

ARTICLE INFO

Keywords:
COVID-19, coronavirus; Pandemic; Viral genomics, blast; SARS-CoV-2

ABSTRACT

The rapid spread of COVID-19, which has led to a global pandemic, has placed public health systems under severe pressure. Identifying variations in SARS-CoV-2 strains from different regions is a key factor for understanding the pathogenic mechanisms, aid in diagnosis, prevention and therapy of this disease. The present study is an analytical descriptive study aimed to determine genetic variations among SARS-CoV-2 strains isolated in China. Sixty six complete genome sequences of the virus were retrieved from NCBI, the sequence of original Wuhan strain accession number NC 045512 was used as the reference sequence. Each genome sequence was blasted against the original Wuhan strain; the analysis was done using NCBI Nucleo-blast. The collected sequences showed 10 different variants. One hundred and thirty four mutations were identified among the variants of SARS-CoV-2 in this study; most of them 52.2% (70/134) were missense point mutation, majority of the mutations 65.7% (88/134) occurred in the open reading frame a/b (ORFab), few mutations occurred in the structural viral genome, each of spike (S) gene and nucleocapsid (N) gene showed 4 mutations; 2 silent point mutations and 2 missense point mutations occurred in each gene whereas membrane (M) gene showed silent point mutation and no mutation observed in the envelope E gene. The remarkable observation in this study showed by Yunnan variant accession number MT226610 which exhibited high incidence of mutations, it displayed 28 different point mutations; only 3 (10.7%) of them were silent mutations while the rest were missense mutations. Our analysis showed several mutations including spike S gene and membrane M gene which may be responsible for a change in the structures of target proteins.

1. Introduction

COVID-19 is one of the most contagious pandemics faced the world, initially it was first reported in December 2019 in Wuhan, China (Abduljalil and Abduljalil, 2020; Deng et al., 2020; Lam, 2020), this pandemic disease had spread to 215 countries and territories around the world, more than 18 million person had infected to date (August 02, 2020), with 689,164 reported deaths, and the cases have increased as high as 5 times in less than a month (worldomater, n.d.). The case numbers under-estimate the real number of infections due to mild or asymptomatic cases and inability to test all population (Deng et al., 2020).

Coronavirus disease (COVID-19) is a respiratory illness primarily affects the respiratory system causing flu-like illness with symptoms such as cough, fever, and in severe cases, difficulty in breathing (Abduljalil and Abduljalil, 2020; Khailany et al., 2020; Naqvi et al., 2020). Because of its marked similarity in terms of biological nature and clinical symptoms with the causative agent of severe acute respiratory syndrome (SARS), the novel coronavirus was termed SARS-CoV-2 by the International Committee on Taxonomy of Viruses (Abduljalil and Abduljalil, 2020; Lokmana et al., 2020). SARS-CoV-2 has been identified as enveloped, positive-sense un-segmented single-stranded RNA viruses that belong to the genus Betacoronavirus, Coronaviridae family of Nidovirales order (Lokmana et al., 2020; Raza et al., 2020; Uddin et al., 2020). The genome size of SARS-CoV-2 approximately 29.9 kb encoding 27 proteins from 14 ORFs including 15 non-structural, 8 accessory, and 4 major structural proteins and lacking the haemagglutinin-esterase gene (Abduljalil and Abduljalil, 2020; Khailany et al., 2020; Lokmana et al., 2020). The longest ORF (ORF1 a/b) is located at the 5' terminus encodes for nonstructural proteins, the 3' terminus of the genome encodes for structural proteins including surface (S), envelope (E), membrane (M), and nucleocapsid (N) proteins with

\textsuperscript{*} Corresponding author at: University of Tabuk, Faculty of Applied Medical Sciences, Department of Medical Laboratory Sciences, Saudi Arabia.

E-mail addresses: eosman@ut.edu.sa (E.H. Ahmed-Abakur), telnour@ut.edu.sa (T.M.S. Alnour).

https://doi.org/10.1016/j.genrep.2020.100925
Received 30 August 2020; Received in revised form 7 October 2020; Accepted 14 October 2020
Available online 17 October 2020
2452-0144/© 2020 Elsevier Inc. All rights reserved.
other accessory proteins (Abduljalil and Abduljalil, 2020; Khailany et al., 2020; Naqvi et al., 2020; Lokmana et al., 2020; Uddin et al., 2020). Based on the phylogenetic studies, the SARS-CoV-2 is categorized in the same lineage that includes SARS coronavirus (SARS-CoV) which causes Severe Acute Respiratory Syndrome (SARS) and MERS-CoV which causes Middle East respiratory syndrome (Lokmana et al., 2020; Raza et al., 2020). The SARS-CoV-2 genome shared about 79–82% sequence identity with MERS-CoV and SARS-CoV (Naqvi et al., 2020; Raza et al., 2020; Uddin et al., 2020). Moreover, the SARS-CoV-2 genome has great sequence similarity (89–96.3%) with two bat coronaviruses; bat-SLCoVZC45 and bat-SL-CoVZXC21 (Abduljalil and Abduljalil, 2020; Uddin et al., 2020).

The World Health Organization (WHO) declared SARS-CoV-2 as a public health emergency of international concern on 30 January 2020 and as a controllable pandemic on 11 March 2020 (World health organization, n.d.). No effective treatment is available for this disease and much is still unknown about it (Abduljalil and Abduljalil, 2020; From the American Association of Neurological Surgeons (AANS) et al., 2018). However, genomic epidemiology of emerging viruses has proven to be a useful tool for outbreak investigation and for tracking virus evolution (Deng et al., 2020). Therefore, the present study aimed to determine the genetic variations in SARS-CoV-2 strains isolated in China.

2. Method and materials

This study is an analytical descriptive study, aimed to determine the genetic variations associated with the SARS-CoV-2 among different strains isolated in China. The study covered 6 Chinese regions (City/province), 67 genome sequences of the virus were retrieved from NCBI Virus Variation Resource repository (https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/) and these strains have been studied according to the time of isolation and the origin. One strain was excluded due to incomplete genomic sequences. The NCBI sequence of SARS-CoV-2 accession number NC045512 which represent original Wuhan strain (submitted on December 2019), was used as the reference sequence. Each genome sequence was blasted against the original Wuhan strain. The analysis was done using NCBI Nucleo-blast and variations of the nucleotides and proteins were reported.

3. Results

Complete genome sequences of the virus from 6 different regions in China (Fig. 1) showed 10 different variants (Table 1). The 1st variant which considered the origin of disease was Wuhan strain which submitten to the gene bank on the mid of December 2019 (accession number NC045512). On the same month similar strains were identified in Beijing (accession numbers MT291827 & MT291830) and in Shanghai (accession number MN908947). The Wuhan strain (variant 1) was continuing to infect human until the end of January 2020.

The sequence identity matrix showed high homology among the viral strains, out of 66 genome sequences of the SARS-CoV-2, 32 (48.5%) were identical (Table 2a) and showed complete genetic similarity to the sequence of the 1st isolate of SARS-CoV-2 (accession number NC045512). Twenty two strains 22 (33.3%) showed 99.99% identity and 10 (15.2%) strains displayed 99.98% identity, the rest 2(3.03%) exhibited less than 99.98% identity to the Wuhan strain NC045512 (Tables 2b, 2c).

One hundred and thirty four mutation were identified among the variants of SARS-CoV-2 in this study; most of them 52.2% (70/134) were missense point mutations, the silent point mutation account for 29.1% (39/134) while non-encoding mutations were 18.7% (25/134), majority of the mutations 65.7% (88/134) occurred in the open reading frame 1a/b (ORF1a/b) which covers most of the viral genome (21,555/29,904 nucleotides), followed by 18.7% (25/134) in 3′UTR terminal loop. All the mutations occurred in the ORF3a, ORF8 and ORF10 were missense point mutations (Fig. 2). Very few mutations observed in structural viral genome; each of spike (S) gene and nucleocapsid (N) gene showed 4 mutations; 1.5% (2/134) silent point mutations and 1.5% (2/134) missense point mutations occurred in each gene, while membrane (M) gene showed only one silent point mutation and no mutation observed in the envelope (E) gene (Fig. 2). The main observation in this study showed in Yunnan variant (accession number MT226610) which exhibited high incidence of mutations, it displayed 28 different point mutations; 1.5% (2/134) silent point mutations and 1.5% (2/134) missense point mutations occurred in each gene, while membrane (M) gene showed only one silent point mutation and no mutation observed in the envelope (E) gene (Fig. 2).

The main observation in this study showed in Yunnan variant (accession number MT226610) which exhibited high incidence of mutations, it displayed 28 different point mutations; only 3(10.7%) of them were silent while the rest were missense mutations. Most of these mutations 26 (89.3%) occurred in the ORF1a/b; one mutation occurred in the S gene (T21784A) and one in ORF8. Another missense point in spike gene mutation was C21711T observed in the Fuyang variant accession number MT281577 (Table 2b).
Table 1

Showed the date of collection, city and main mutations of variants of SARS-CoV-2 which specified in this study.

| December 2019 | January 2020 | February | March |
|---------------|-------------|----------|-------|
| Wuhan         | NC_045512 (V1) | Wuhan    | Wuhan |
| MT059631      | LR757996    | MT059631 | Shanghai MT121215 |
| MT059632      | MN908668    | MN996527 | C6822T (M) |
| MN908669      | MN988668    | MT019530 | C12475T (S) |
| Beijing       | Beijing     | Beijing  | Beijing |
| MT291827      | MT291828    | MT291835 | MT291834 |
| MT291830      | MT291829    | MT291835 | MT291834 |
| Wuhan         | Wuhan      | Wuhan   | Wuhan |
| LR757998      | MT019531    | LR757996 | MT019529 |
| C9068A (S),   | Wuhan     | T721C (S) | G6819T (M) |
| T11764A (S)   | Wuhan     | G1895T (M) | G11083T (M) |
| Wuhan         | Beijing    | Beijing  | Beijing |
| MT291826      | MT291828    | MT291832 | MT291833 |
| C972T (S)     | C972T (S)  | C972T (S) | C972T (S) |
| T28144C (M)   | T28144C (S) | T28144C (S) | T28144C (S) |
| T28144C (M)   | T28144C (S) | T28144C (S) | T28144C (S) |
| Guangdong     | Guangdong  | Guangdong | Guangdong |
| MT18334       | MT18334    | MT18334 | MT18334 |
| C972T (S)     | C972T (S)  | C972T (S) | C972T (S) |
| T28144C (M)   | T28144C (S) | T28144C (S) | T28144C (S) |
| T28144C (M)   | T28144C (S) | T28144C (S) | T28144C (S) |
| Yunnan        | Yunnan     | Yunnan   | Yunnan |
| MT049951      | MT049951   | MT049951 | MT049951 |
| C972T (S)     | C972T (S)  | C972T (S) | C972T (S) |
| T28144C (M)   | T28144C (S) | T28144C (S) | T28144C (S) |
| T28144C (M)   | T28144C (S) | T28144C (S) | T28144C (S) |
| Guangdong     | Guangdong  | Guangdong | Guangdong |
| MT123291      | MT123291   | MT123291 | MT123291 |
| C20692T (M)   | C20692T (S) | C20692T (S) | C20692T (S) |
| C29868C (M)   | C29868C (S) | C29868C (S) | C29868C (S) |
| Wuhan         | Wuhan     | Wuhan   | Wuhan |
| MT291827      | MT291828    | MT291827 | MT291827 |
| MT291829      | MT291828    | MT291827 | MT291827 |
| MT291829      | MT291828    | T1623C (M) | T1623C (M) |
| G29868C (M)   | T29846A (M) | T29846A (M) | T29846A (M) |
| T29847A (M)   | T29847A (M) | T29847A (M) | T29847A (M) |
| T29847A (M)   | T29847A (M) | T29847A (M) | T29847A (M) |
| G29868C (M)   | G29868C (M) | G29868C (M) | G29868C (M) |
| Guangdong     | Guangdong  | Guangdong | Guangdong |
| MT123291      | MT123291   | MT123291 | MT123291 |
| C17373T (S)   | C17373T (S) | C17373T (S) | C17373T (S) |
| G29868C (M)   | G29868C (M) | G29868C (M) | G29868C (M) |
| Guangdong     | Guangdong  | Guangdong | Guangdong |
| MT123291      | MT123291   | MT123291 | MT123291 |
| C17373T (S)   | C17373T (S) | C17373T (S) | C17373T (S) |
| G29868C (M)   | G29868C (M) | G29868C (M) | G29868C (M) |
| Guangdong     | Guangdong  | Guangdong | Guangdong |
| MT123291      | MT123291   | MT123291 | MT123291 |
| C17373T (S)   | C17373T (S) | C17373T (S) | C17373T (S) |
| G29868C (M)   | G29868C (M) | G29868C (M) | G29868C (M) |

E.H. Ahmed-Abakur and T.M.S. Alnour
Gene Reports 21 (2020) 100925
E.H. Ahmed-Abakur and T.M.S. Alnour

Variants 2–6 (Table 1) sequenced during December 2019 were detected in all Chinese regions except Hangzhou, these variants characterized by missense point mutation in the Open Reading Frame 1 A and B (ORF1 a/b). Variant 2 (accession number LR757996) was observed later in Beijing, Guangdong and Yunnan. This variant showed C8782T silent point mutation in ORF1 a/b and T28144C missense mutation in the ORF8, upon spread in various regions extra mutations were observed in Beijing (accession numbers MT291826, MT291831-34, MT135043); Guangdong (accession numbers MN938384, MN975262 and MT123292); and in Yunnan (accession numbers MT049951 and MT226610), C29095T is silent mutation of the N gene observed in the Guangdong variants (accession numbers MN938384 and MN975262). Silent point mutation T21644A in S gene was observed in Yunnan variants (accession number MT049951) (Table 2c).

Table 2a
showed the completely identical sequences of SARS CoV-2 variants compared to Wuhan variant NC045512.

| Number | City     | Accession # | Collection date | % of identity to Wuhan |
|--------|----------|-------------|-----------------|------------------------|
| 1      | Wuhan    | NC_045512   | December 2019   | 100%                   |
| 2      | Beijing  | MT093631    | 8-1-2020        | 100%                   |
| 3      | Beijing  | MT039874    | 22-1-2020       | 100%                   |
| 4      | Wuhan    | MN988669    | 2-1-2020        | 100%                   |
| 5      | Wuhan    | MN988668    | 2-1-2020        | 100%                   |
| 6      | Shanghai | MN098947    | 12-2019         | 100%                   |
| 7      | Hangzhou | MT039873    | 20-1-2020       | 100%                   |
| 8      | Wuhan    | MT019533    | 1-1-2020        | 100%                   |
| 9      | Wuhan    | MT019532    | 30-12-2019      | 100%                   |
| 10     | Wuhan    | MT019531    | 30-12-2019      | 100%                   |
| 11     | Beijing  | MT291830    | 30-12-2019      | 100%                   |
| 12     | Beijing  | MT291829    | 30-12-2019      | 100%                   |
| 13     | Beijing  | MT291828    | 30-12-2019      | 100%                   |
| 14     | Beijing  | MT291827    | 30-12-2019      | 100%                   |
| 15     | Wuhan    | MN996530    | 30-12-2019      | 100%                   |
| 16     | Wuhan    | MN996528    | 30-12-2019      | 100%                   |
| 17     | Hangzhou | MT253710    | 21-1-2020       | 100%                   |
| 18     | Hangzhou | MT253709    | 21-1-2020       | 100%                   |
| 19     | Hangzhou | MT253708    | 21-1-2020       | 100%                   |
| 20     | Hangzhou | MT253707    | 25-1-2020       | 100%                   |
| 21     | Hangzhou | MT253706    | 22-1-2020       | 100%                   |
| 22     | Hangzhou | MT253705    | 22-1-2020       | 100%                   |
| 23     | Hangzhou | MT253704    | 25-1-2020       | 100%                   |
| 24     | Hangzhou | MT253703    | 25-1-2020       | 100%                   |
| 25     | Hangzhou | MT253702    | 21-1-2020       | 100%                   |
| 26     | Hangzhou | MT253701    | 21-1-2020       | 100%                   |
| 27     | Hangzhou | MT253700    | 25-1-2020       | 100%                   |
| 28     | Hangzhou | MT253699    | 24-1-2020       | 100%                   |
| 29     | Hangzhou | MT253698    | 23-1-2020       | 100%                   |
| 30     | Hangzhou | MT253697    | 23-1-2020       | 100%                   |
| 31     | Hangzhou | MT253696    | 23-1-2020       | 100%                   |
| 32     | Wuhan    | LR757996    | 1-1-2020        | 100%                   |

Table 2b
showed the mutations in SARS CoV-2 Yunnan variant MT226610 compared to Wuhan variant NC045512.

| Number | City   | Accession # | Collection date | % of identity to Wuhan | Mutations locations | Target gene | Protein |
|--------|--------|-------------|-----------------|------------------------|---------------------|-------------|---------|
| 33     | Yunnan | MT226610    | 20-1-2020       | 99.91%                 | G428T    ORF1ab    ORF1a/b | E1341D     |
|        |        |             |                 |                        | A4307C    ORF1a/b   ORF1a/b | K1348Q     |
|        |        |             |                 |                        | A7479G    ORF1a/b   ORF1a/b | N2405S     |
|        |        |             |                 |                        | C8782T    ORF1a/b   ORF1a/b | Sil M      |
|        |        |             |                 |                        | G1108T    ORF1a/b   ORF1a/b | L3606F     |
|        |        |             |                 |                        | G11207C   ORF1a/b   ORF1a/b | A3648P     |
|        |        |             |                 |                        | T11233G   ORF1a/b   ORF1a/b | Silent M   |
|        |        |             |                 |                        | G12041C   ORF1a/b   ORF1a/b | E3965D     |
|        |        |             |                 |                        | G12160C   ORF1a/b   ORF1a/b | K3079N     |
|        |        |             |                 |                        | G12202C   ORF1a/b   ORF1a/b | K3981N     |
|        |        |             |                 |                        | G12206T   ORF1a/b   ORF1a/b | Q4030H     |
|        |        |             |                 |                        | G12355C   ORF1a/b   ORF1a/b | R4038K     |
|        |        |             |                 |                        | G12378A   ORF1a/b   ORF1a/b | A4067S     |
|        |        |             |                 |                        | G12464T   ORF1a/b   ORF1a/b | A4068S     |
|        |        |             |                 |                        | G12467T   ORF1a/b   ORF1a/b | D4076Y     |
|        |        |             |                 |                        | G12491T   ORF1a/b   ORF1a/b | Silent M   |
|        |        |             |                 |                        | G12514C   ORF1a/b   ORF1a/b | D4103Y     |
|        |        |             |                 |                        | G12572T   ORF1a/b   ORF1a/b | D4105Y     |
|        |        |             |                 |                        | G12578T   ORF1a/b   ORF1a/b | S4106L     |
|        |        |             |                 |                        | G12658T   ORF1a/b   ORF1a/b | S4112N     |
|        |        |             |                 |                        | G12660A   ORF1a/b   ORF1a/b | R4132T     |
|        |        |             |                 |                        | G12660C   ORF1a/b   ORF1a/b | Q4140H     |
|        |        |             |                 |                        | G12685C   ORF1a/b   ORF1a/b | A4170S     |
|        |        |             |                 |                        | G12773T   ORF1a/b   ORF1a/b | K4176N     |
|        |        |             |                 |                        | G12795T   ORF1a/b   ORF1a/b | D6906H     |
|        |        |             |                 |                        | G20980C   ORF1a/b   ORF1a/b | N6997K     |
|        |        |             |                 |                        | T21784A   ORF1a/b   ORF1a/b | L9117S     |
|        |        |             |                 |                        | T28144C   ORF1a/b   ORF1a/b |  **M**    |

4. Discussion

Identifying variations in strains from different regions is a key factor for understanding the pathogenic mechanisms of this disease (From the
| Number | City                  | Accession #          | Collection date | % of identity to Wuhan | Mutations locations | Target gene | Protein                        |
|--------|-----------------------|----------------------|-----------------|------------------------|--------------------|-------------|--------------------------------|
| 34     | Yunnan Province       | MT396241              | 6-3-2020        | 99.99%                 | G1433A             | ORF1ab      | E390K                          |
| 35     | Shenzhen, Guangdong   | MN975262              | 11-1-2020       | 99.98%                 | G0561T             | ORF1ab      | D5216Y                         |
| 36     | Fuyang, Anhui         | MT281577              | 10-3-2020       | 99.98%                 | T28144C            | ORF1b       | S3099L                         |
| 37     | Beijing               | MT291835              | 27-1-2020       | 99.99%                 | T7077C             | ORF1b       | L9117S                         |
| 38     | Beijing               | MT291834              | 28-1-2020       | 99.98%                 | G2867A             | Silent      | L9117S                         |
| 39     | Beijing               | MT291833              | 28-1-2020       | 99.99%                 | T4402C             | ORF1b       | Silent                          |
| 40     | Beijing               | MT291832              | 25-1-2020       | 99.99%                 | G5062T             | ORF1b       | L9117S                         |
| 41     | Beijing               | MT135044              | 28-1-2020       | 99.99%                 | C8782T             | ORF1b       | Silent                          |
| 42     | Beijing               | MT135042              | 28-1-2020       | 99.99%                 | T28144C            | ORF1b       | L9117S                         |
| 43     | Beijing               | MT135041              | 28-1-2020       | 99.99%                 | A29695G            | ORF1b       | D9503V                         |
| 44     | Beijing               | MT291831              | 24-1-2020       | 99.98%                 | C18661G            | Silent      | P9504S                         |
| 45     | Beijing               | MT135043              | 28-1-2020       | 99.98%                 | T4402C             | Silent      | C3891Y                         |
| 46     | Guangdong, Guangzhou  | MT123290              | 5-2-2020        | 99.99%                 | G564A              | ORF1b       | L9117S                         |
| 47     | Guangdong, Guangzhou  | MT123293              | 29-1-2020       | 99.98%                 | G654A              | Silent      | L9117S                         |
| 48     | Guangdong, Guangzhou  | MT123292              | 27-1-2020       | 99.99%                 | G654A              | Silent      | L9117S                         |
| 49     | Guangdong, Guangzhou  | MT123291              | 29-1-2020       | 99.99%                 | G654A              | Silent      | L9117S                         |
| 50     | Wuhan                 | MT259231              | 25-1-2020       | 99.99%                 | G29868C            | 3'UTR       | G130E                          |
| 51     | Wuhan                 | MT259229              | 26-1-2020       | 99.99%                 | G29868C            | 3'UTR       | S2185I                         |
| 52     | Wuhan                 | MT259230              | 25-1-2020       | 99.99%                 | G29868C            | 3'UTR       | L2244T                         |
| 53     | Wuhan                 | MT259228              | 26-1-2020       | 99.98%                 | G29868C            | 3'UTR       | P68105                         |
| 54     | Wuhan                 | MT259227              | 26-1-2020       | 99.97%                 | G29868C            | 3'UTR       | S6973G                         |
| 55     | Wuhan                 | MT259226              | 10-1-2020       | 99.98%                 | G29868C            | 3'UTR       | T8553I                         |

(continued on next page)
The present study exhibited 10 different variants; this finding was in alignment with several reports which showed several variants of SARS-CoV-2 (Abduljalil and Abduljalil, 2020; Uddin et al., 2020; Yu et al., 2020). Forster et al., carried out study using a phylogenetic network analysis approach on 160 full length genomes and reported that the virus seems to be evolving into three central variants distinguished by amino acid changes (Forster et al., 2020). Based on phylogenetic analyses Abduljalil and Abduljalil concluded that SARS-CoV-2 genomes sequenced showed genomic variations among strains from different regions and countries through different period (Abduljalil and Abduljalil, 2020). This diversity may be owing to the nature of virus itself where the RNA viruses tend to harbor error-prone RNA dependent RNA poly-merases which makes occurrence of mutations and recombination events rather frequently (Uddin et al., 2020). Kupferschmidt stated that SARS-CoV-2 like other coronaviruses appears to accumulate, on average, one or two mutations per month (Kupferschmidt, 2020).

Our study showed that the majority of the mutations 65.7% (88/134)
occurred in the open reading frame 1 a/b (ORF1 a/b), similar result concerning the frequency of mutations in ORF1 a/b was reported by Khailany et al. (2020) ORF1 a/b covers most of the viral genome and encoded for nonstructural proteins (Abduljalil and Abduljalil, 2020; Lokmana et al., 2020) that collectively involved in virus replication and possibly in immune evasion (Abduljalil and Abduljalil, 2020). Hurst et al. (2013) reported that there was a basic connection between the nsp3 association and the inception of coronavirus infection whereas Wan et al. mentioned that the area of ORF1a/b is the most important factor among coronaviruses (Wan et al., 2020).

Spike protein (S) of SARS coronavirus (SARS-CoV) play a major role in SARS-CoV-2 pathogenesis, it attaches the virus to its cellular receptor, angiotensin-converting enzyme 2 (ACE2) (World health organization, n.d.; Wan et al., 2020). Our study showed limited mutations in the structural genes, each of spike (S) gene and nucleocapsid (N) gene revealed 4 mutations; 2 silent point mutations and 2 missense point mutations, while membrane (M) gene showed only one silent point mutation and no mutation observed in the envelope (E) gene, this findings are in alignment with other study conducted by Lokmana et al. who analyzed 320 whole-genome sequences and 320 spike protein sequences of SARS-CoV-2 and reported just one deletion in the spike (S) protein (Lokmana et al., 2020) and with Wang et al. who studied genetic variations among 95 full length genomic sequences of SARS-CoV-2 strains and they stated that SARS-CoV-2 is relatively conserved, especially in the E region (Wang et al., 2020). The viral spike protein is thought to have a crucial role in drug and vaccine development as reported previously in managing the viruses like SARS-CoV and MERS-CoV (Tian et al., 2020), therefore mutation in this gene might affect the severity and spreads of the SARS-CoV-2 as well suppresses the efforts of developing vaccine.

In this study the Yunnan variant accession number MT226610 exhibited high incidence of mutations; it displayed 28 different point mutations including one mutation in the S gene (T21789A), another missense point in spike gene mutation was C21711T observed in the Fuyang variant accession number MT281577. The T21784A point mutation can lead to replaced asparagine by leucin whereas the MT281577 point mutation leads to substitute serine with leucin in the position 9693 of the amino acid sequence, this altering might affect the viral protein functions. However, hyper mutations may change the virus proteins, thus affecting the virus behavior and potency; this may lead to different waves of disease and complicated the efforts of producing vaccines.

Several common gene mutations were observed among the SARS-CoV-2 sequence in China. These mutations follow standard roles and common among different countries. The most counted mutations in our study were T4402C, G5062T, C8782T, C17373T, C20692T, T28144C, C29095T, and G29886C. The T4402C mutation which leads to silent mutation in the ORF1 a/b gene segment was noted in the strain isolated from Beijing (variant # 2) and always associated with C8782T, G5062T and T28144C mutations, similar T4402C and G5062T point mutation were observed in two strains isolated in South Korean (Yang et al., 2020), C8782T was the predominant mutation reported in SARS-CoV-2 gene mutation around the world (Yang et al., 2020; Mercatelli and Giorgi, 2020), this mutation is always coexisting with the missense point mutation of the ORF8 gene segment T28144C (Mercatelli and Giorgi, 2020). The C17373T silent mutation was observed in Wuhan and tends to spread to Guangdong. Same mutation had noticed in Singapore and USA (Li et al., 2020). The C20692T was restricted to Wuhan and was coexisting with the G29886C gene mutation of the 3 terminal loop. This mutation was also noticed by Sharp and Dange (2020). The C29095T mutation of the N gene was also reported in the USA (Yang et al., 2020).

5. Conclusion

Study the sequencing of coronavirus at different points/regions can tell how the virus is adapting and can indicate the disease epidemiology, pathogenesis and may help in developing treatment and prevention. Our analysis revealed several mutations including spike S gene and membrane M gene which may be responsible for a change in the structures of target proteins.

Declaration of competing interest

This research is NOT funded fully or partially from any governmental or non-governmental agencies.

References

Abduljalil, J.M., Abduljalil, B.M., 2020. Epidemiology, genome, and clinical features of the pandemic SARS-CoV-2: a recent view. New Microbes New Infect. 35, 100672. Published 2020 Mar 31. https://doi.org/10.1016/j.mni.2020.100672.

Deng, X., et al., 2020. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into northern California. Science. https://doi.org/10.1126/science.abc2963.

Forster, P., Forster, L., Renfrew, C., Forster, M., 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc. Natl. Acad. Sci. U. S. A. 117, 9241–9243.

From the American Association of Neurological Surgeons (AANS), American Society of Neuroradiology (ASNR), Cardiovascular and Interventional Radiology Society of Europe (CIRSE), Canadian Interventional Radiology Association (CIRA), Congress of Neurological Surgeons (CNS), European Society of Minimally Invasive Neurological Therapy (ESMINT), European Society of Neuroradiology (ESNR), European Stroke Organisation (ESO), Society for Cardiovascular Angiography and Interventions (SCAI), Society of Interventional Radiology (SIR), Society of Neuroradiology (SNS), World Stroke Organization (WSO), Sacks, D., Baxter, B., et al., 2018. Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. Int. J Stroke 13 (6), 612–622. https://doi.org/10.1177/1747493018778713.

Hurst, K.R., Koetzner, C.A., Masters, P.S., 2013. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J. Virol. 87 (6), 9159-9175.

Khailany, R.A., Sa’dar, M., Ozaslan, M., 2020. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 19, 100682.

Kupferschmidt, K., 2020. Genome analyses help track coronavirus’ moves. Science 367 (6485), 1176e1177.

Lam, T.T., 2020. Tracking the genomic footprints of SARS-CoV-2 transmission. Trends Genet. 36 (8).

Li, J., Li, Z., Cai, X., Wu, C., 2020. Bayesian phylogenetic inference on the temporal evolution and global transmission of SARS-CoV-2. J. Inf. Secur. 81 (2), 318–356. Aug.

Lokmana, S.M., Rashaduzzaman, Saluddin, A., Barua, R., Tanzina, A.Y., Rumi, M.H., Hossain, I., Siddiki, Z., Mannan, A., Hasan, M., 2020. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. Infect. Genet. Evol. 84, 104389.

Mercatelli, D., Giorgi, F., 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. Front. Microbiol. 11, 1800. https://doi.org/10.3389/fmicb.2020.01800.

Naqui, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, L.K., Singh, A., Atif, S.M., Harirpadaad, G., Hasan, G.M., Hassan, L., 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapy: a structural genomics approach. Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1866, 165878.

Raza, H., Wahid, B., Rubih, G., Gulzar, A., 2020. Molecular epidemiology of SARS-CoV-2 in Faisalabad, Pakistan: a real world clinical experience. Infect. Genet. Evol. 84, 104074.

Sharq, K., Dange, S., 2020. May. Comparative analysis of SARS-CoV2 with special emphasis of the genome sequences. J. Pure Appl. Microbiol. 14 (suppl1), 1053-1064. https://doi.org/10.22207/JPAM.14.SP1.44.

Tian, X., Li, C., Huang, A., et al., 2020. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg. Microbes Infect. 9 (1), 382–385.

Uddin, M., Mustafa, F., Rizvi, T.A., Loney, T., Al Suwaidi, H., Al-Marzouqi, A.H.H., Eldin, A.K., Abuzeeba, N., Adrian, T.E., Stefani, C., Nowotny, N., Alsheikh-Ali, A., Senek, A.C., 2020. SARS-CoV-2/COVID-19: viral genomics, epidemiology, vaccines, and therapeutic interventions. Viruses 12, 526. https://doi.org/10.3390/v12040526.

Wan, Y., Shang, J., Graham, R., Baric, R.S., Li, F., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J. Virol. 94, 4. https://doi.org/10.1128/JVI.01127-20 (e01127-20).

Wang, C., Liu, Z., Chen, Z., Huang, X., Yu, M., He, T., Zhang, Z., 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. J. Med. Virol. https://doi.org/10.1002/jmv.25762.

World health organization. https://www.who.int/news-room/detail/29-06-2020-covid19.

Yu, W., Tang, G., Zhang, L., Corlett, R., 2020. Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2/2019-nCoV) using whole genomic data. Zool. Res. 41 (3), 247e257.