



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.



Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Phylogenetic and amino acid signature analysis of the SARS-CoV-2s lineages circulating in Tunisia

Mouna Ben Sassi^{a,b,1}, Sana Ferjani^{c,d,*}, Imen Mkada^{e,f,1}, Marwa Arbi^{g,1}, Mouna Safer^h, Awatef Elmoussi^c, Salma Abid^c, Oussema Souiai^g, Alya Gharbi^{e,f,i}, Asma Tejouri^j, Emna Gaies^{a,b}, Hanene Eljabri^a, Samia Ayed^{i,k}, Aicha Hechaichi^h, Riadh Daghfous^{a,b}, Riadh Gouider^{e,f,i}, Jalila Ben Khelil^{i,k}, Maher Kharrat^j, Imen Kacem^{e,f,i}, Nissaf Ben Alya^h, Alia Benkahla^g, Sameh Trabelsi^{a,b}, Ilhem Boutiba-Ben Boubaker^{c,d}

^a National Center Chabibekahia of Pharmacovigilance of Tunis, Laboratory of Clinical Pharmacology, Boulevard 9 Dr Zouhier Essafi, Tunis 1006, Tunisia

^b University of Tunis El Manar, Faculty of Medicine of Tunis, LR16SP02, 1007 Tunis, Tunisia

^c Charles Nicolle Hospital, Laboratory of Microbiology, Virology unit, National Influenza and other Respiratory Viruses Center-Tunisia, Boulevard 9 Avril, Tunis 1006, Tunisia

^d University of Tunis El Manar, Faculty of Medicine of Tunis, LR99ES09, 1007 Tunis, Tunisia

^e Department of Neurology, LR18SP03, Razi University Hospital, 1 rue des oranges Manouba, 2010 Tunis, Tunisia

^f Clinical Investigation Centre (CIC) "Neurosciences and Mental Health", Razi University Hospital, 1 rue des oranges Manouba, 2010, Tunis, Tunisia

^g Laboratory of Bioinformatics, Biomathematics and Biostatistics-LR16IPT09, Institut Pasteur de Tunis, University of Tunis El Manar (UTM), Tunis, Tunisia

^h Ministry of Health, National Observatory of New and Emerging Diseases, 1006 Tunis, Tunisia

ⁱ Faculty of Medicine of Tunis, University of Tunis El Manar, 15, Rue DjebelAkhdar - La Rabta - 1007, Tunis, Tunisia

^j University of Tunis El Manar, Faculty of Medicine of Tunis, LR99ES10 Human Genetics Laboratory, 1007 Tunis, Tunisia

^k Abderrahmen Mami Hospital, Department of Medical Intensive Care, 2080 Ariana, Tunisia

ARTICLE INFO

Keywords:

Coronavirus disease-2019

SARS-CoV-2

Whole genome sequencing

Lineages phylogenetic

Amino acid change analysis

Amino acid signature

Tunisia

ABSTRACT

Since the beginning of the Coronavirus disease-2019 pandemic, there has been a growing interest in exploring SARS-CoV-2 genetic variation to understand the origin and spread of the pandemic, improve diagnostic methods and develop the appropriate vaccines. The objective of this study was to identify the SARS-CoV-2s lineages circulating in Tunisia and to explore their amino acid signature in order to follow their genome dynamics. Whole genome sequencing and genetic analyses of fifty-eight SARS-CoV-2 samples collected during one-year between March 2020 and March 2021 from the National Influenza Center were performed using three sampling strategies. Multiple lineage introductions were noted during the initial phase of the pandemic, including B.4, B.1.1, B.1.428.2, B.1.540 and B.1.1.189. Subsequently, lineages B1.160 (24.2%) and B1.177 (22.4%) were dominant throughout the year. The Alpha variant (B.1.1.7 lineage) was identified in February 2021 and firstly observed in the center of our country. In addition, A clear diversity of lineages was observed in the North of the country. A total of 335 mutations including 10 deletions were found. The SARS-CoV-2 proteins ORF1ab, Spike, ORF3a, and Nucleocapsid were observed as mutation hotspots with a mutation frequency exceeding 20%. The 2 most frequent mutations, D614G in S protein and P314L in Nsp12 appeared simultaneously and are often associated with increased viral infectivity. Interestingly, deletions in coding regions causing consequent deletions of amino acids and frame shifts were identified in NSP3, NSP6, S, E, ORF7a, ORF8 and N proteins.

These findings contribute to define the COVID-19 outbreak in Tunisia. Despite the country's limited resources, surveillance of SARS-CoV-2 genomic variation should be continued to control the occurrence of new variants.

* Corresponding author at: Laboratory of Microbiology, Charles Nicolle Hospital, Boulevard du 9 avril 1938, 1006 Tunis, Tunisia.

E-mail address: ferjsana@yahoo.fr (S. Ferjani).

¹ The first four authors have equally contributed in this study.

<https://doi.org/10.1016/j.meegid.2022.105300>

Received 8 September 2021; Received in revised form 1 May 2022; Accepted 5 May 2022

Available online 10 May 2022

1567-1348/© 2022 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Coronavirus disease-2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a growing public health concern. In some people, produces an asymptomatic disease or mild symptoms disease that does not require particular medical care. However, in specific groups of patients, particularly the elderly and those with chronic health diseases, the infection progresses into severe respiratory distress, requiring hospitalization in intensive care units (Thielen et al., 2021).

SARS-CoV-2 genome like other RNA viruses, shows a high mutation rates. Initially, the virus emerged from an animal reservoir in the city of Wuhan, China. Then, a human-to-human transmission with a rapid spread worldwide has been established (Chunyang et al., 2020). Over one year of COVID-19 pandemic, new SARS-CoV-2 genome mutations were constantly emerging and more than 4000 variants have been reported (Bian et al., 2021). Epidemiological and phylogenetic studies revealed that the significant increase in the rate of infection and/or death was correlated with the emergence of four Variant of Concern (VOC) belonging to B.1.1.7, B.1.351, B.1.617.2 and B.1.1.28 lineages. On May 31, 2021, WHO published a new nomenclature proposal applying to VOC based on the Greek alphabet: Alpha, Beta, Delta and Gamma, respectively (Parums, 2021).

During the initial stage of the pandemic, due to the lack of specific treatments that prevent or block viral replication, massive prevention strategies were applied by most countries. A wide difference in case fatality rates was observed, probably due to a diverse demographic composition and the type of measures that were taken in different countries to limit viral spread (Rader et al., 2021). Subsequently, following the widespread vaccination, SARS-CoV-2 infections and deaths declined and social and economic conditions improved relatively. Globally, 196.553.009 confirmed cases of COVID-19 have been identified, with 4.200.412 deaths as of 30 of July 2021 (World Health Organization (WHO), 2021a).

In Tunisia, the first case of COVID-19 was identified on 3 March 2020. Preventive strategies were quickly put in place, in particular, lockdown and enhanced contact tracing around all positive cases (Chakroun et al., 2020). As of 25 May 2020, the cumulative number of confirmed cases of COVID-19 was 1051 corresponding to a cumulative incidence of 8.87/100,000 inhabitants and an average of daily incidence around 13 cases (Abid et al., 2020). In view of the critical socio-economic situation, the Tunisian authorities allowed the reduction of restrictions. Accordingly, the virus has continued to spread with alarming rates, recording 595.532 positive cases and 20.067 deaths on 30th July 2021 (World Health Organization (WHO), 2021a). The low vaccine administration rate (11%) coupled with the emergence and wide circulation of the different Variants of Concern (VOC) had certainly played an important role in the evolution of the epidemiological situation in Tunisia (World Health Organization (WHO), 2021b).

The aim of this study was to identify the different SARS-CoV-2 introduction events in Tunisia, and to explore mutation profile to follow their genome dynamics through a collection of SARS-CoV-2 strains ($n = 58$) from National Influenza Center over one year of the COVID-19 pandemic (March 2020 to March 2021).

2. Materials and methods

2.1. Ethical statement

The PRFCOVID-GP3 project titled “SARS-CoV-2 genome Sequencing and study of host-pathogen molecular interactions in Tunisia: epidemiological, clinical and therapeutic impact” was approved by the medical ethics committee of Razi Hospital of Tunis. All procedures involving human participants were in accordance with the ethical standards of the Medical Ethics Committee of Razi Hospital and with the 1964 Helsinki declaration.

2.2. Sampling strategies

In total 90 SARS-CoV-2 strains resulting from three sampling strategies during one year of the COVID-19 pandemic were included in this study. Initially, a stratified random sampling ($n = 32$) was performed from February 5 to July 17, 2020 (first phase of the epidemic, well-controlled) according to the following criteria: super spread events ($n = 1$), extreme evolution ($n = 3$), death ($n = 3$), local infection ($n = 12$) and imported infection ($n = 13$) [France ($n = 4$), Italy ($n = 1$), England ($n = 1$), Egypt ($n = 1$), Switzerland ($n = 1$), Turkey ($n = 4$) and Spain ($n = 1$)]. Then from July 18 to December 23, 2020, a simple random sampling was applied from the list of samples with positive RT-PCR ($n = 32$). Finally, from December 24 to March 20, 2021, the sequencing indications were in accordance with the national sequencing strategy of SARS-CoV-2. It aimed to identify and monitor the VOCs emergence in Tunisia ($n = 26$).

Thirty-two samples were excluded, due to high Ct value ($n = 10$), amplification failure during sequence processing ($n = 17$) or poor genomic coverage ($<60\%$, $n = 5$). Thus, 58 samples were included in this study (Supplementary Data Table S1). None of the included samples were collected in May and June 2020, due to the absence of COVID-19 cases in Tunisia.

2.3. Viral RNA extraction and Real-time RT-PCR

All nasopharyngeal samples included in this study were collected at the National Influenza Center (NIC), also nominated as National Reference Lab for SARS-CoV-2 and other Respiratory Viruses and hosted at the Microbiology lab of Charles Nicolle Hospital of Tunis. RNA extraction was performed using the Chemagic™ automate and the viral RNA 300 Kit H96 (Perkin Elmer, Hamburg, Germany) according to the manufacturer's instructions. Also, manual extraction using the Qiagen Viral RNA Mini Kit (QIAGEN, California, USA) was used depending on the availability of reagents. SARS-CoV-2 was detected by the Hong Kong RT-PCR assay using AgPath-ID™ one-Step RT-PCR Reagents and ABI 7500 instrument (WHO, Laboratory and diagnosis, 2020). Which is a qualitative real time RT-PCR TaqMan method. According to this assay, a positive COVID-19 result was determined when both targets N and ORF1b-nsp14 reach a defined threshold below 0.2 and Ct value below 40.

2.4. Library preparation, sequencing, and read processing

All samples included in this study had a Ct value <30 . WGS of SARS-CoV-2 strains were performed using the Illumina® RNA Prep with Enrichment with Illumina Respiratory Virus Oligos Panel at the National Center of Pharmacovigilance Chalbi Belkahia of Tunis. Initially, RNA was quantified using Spectrophotometer/Fluorometer DeNovix® DS-11 series. Input amounts of Illumina® RNA Prep ranged from 10 to 100 ng of RNA. Tiling Polymerase Chain Reaction cDNA synthesis was performed using the Illumina cDNA Synthesis kit. The cDNA was quantified using dsDNA High Sensitivity DeNovix® kit. To tag the cDNA, Enrichment Bead-Linked Transposomes were used. After tagging, the fragments were purified and amplified to add index adapter sequences for dual indexing. Then, samples were enriched as single-plex reactions using the Respiratory Virus Oligos Panel v2 (Illumina, Catalog no. 20044311), which features ~7800 probes designed to detect respiratory viruses, recent flu strains, and SARS-CoV-2, as well as human probes to act as positive controls in each reaction. Libraries were run on an Illumina iSeq100™. The prepared libraries were diluted to a final loading concentration of 100 pM, according to the iSeq 100 System denature and dilute Libraries Guide and sequenced on the iSeq 100 System at 2×150 bp read length. FASTQ sequencing data files were input to the DRAGEN RNA Pathogen Detection pipeline® and the ID by DNAXplify Respiratory Virus Oligos Panel Platform® for analysis and viral detection. These platforms were accessed in BaseSpace Sequence Hub.

2.5. Lineage assignment

Clades were assigned to SARS-CoV-2 genome sequences ($n = 58$) using the Phylogenetic Assignment of Named Global Outbreak LINEages tool (PANGOLIN), NextStrain and GISAID nomenclature systems. These are updated based on newly observed viral lineages. PANGOLIN available at <https://cov-lineages.org>, currently recognizes two lineages: A and B. Lineage A genomes are characterized by two unique mutations (8782C > T and 28,144 T > C), compared to lineage B. From these lineages, sub-lineages (e.g. A.1, A.2, A.3, ...) and sub sub-lineages (e.g. A.1.1) are designated, each defined by additional mutations and specific epidemiological characteristics (Rambaut et al., 2020). Next Strain clade designations, (<https://clades.nextstrain.org/>), separate viruses that originated in China in 2019 (Clade 19) from those that subsequently introduced into Europe in early 2020 (Clade 20). There were two clades identified in 2019 (19A and 19B) and 9 more in 2020 (20A to 20I). Subclades within a major clade were designated by specific nucleotide mutations. GISAID (<https://www.gisaid.org/>) uses specific combinations of genetic markers. Currently eight clades are defined: S and L, to the further evolution of L into V and G, and later of G into GH, GR and GV, and more recently GR into GRY. The lineages assigned by the PANGOLIN nomenclature system were used to discuss viral diversity throughout this manuscript (updated on august 29, 2021).

2.6. Phylogenetic reconstructions and analysis

A phylogenetic tree was built from the 58 full-length Tunisian sequences and the reference NC_045512 sequence using approximate maximum-likelihood (ML) method of MEGA X software (Sneath and Sokal, 1973) based on 1000 bootstrap replicates.

2.7. Genomic analysis

The 58 SARS-CoV-2 genome sequences were aligned using Clustal W program (Larkin et al., 2007) implemented in MEGA X software (Kumar et al., 2018). Multiple alignments were manually edited by trimming the 5' and 3' untranslated regions, removing gaps and low-quality

sequences, and then visualized using MEGA X. In addition, Open Reading Frames (ORFs) were predicted and annotated following the annotation of the SARS-CoV-2 reference genome generated from the Wuhan-Hu-1 sequence (access number: NC_045512). Each genome was compared to the reference NC_045512, then, genomic variants were identified using Geneious software (Kearse et al., 2012). Frequencies of identified variants were calculated and plotted according to their position on NC_045512 using Graph Pad Prism v8 (Graph Pad Software, Inc. San Diego California, USA). Mutations with frequencies above 20% were considered as hot spots.

3. Results

3.1. Lineage analysis

A total of 15 different lineages were identified among the 58 SARS-CoV-2 genome sequences generated in this study. Most sequences belonged to lineage B1.160 ($n = 14$; 24.1%) and B1.177 ($n = 13$; 22.4%). Most health care workers included in this study ($n = 8$) carried B1.160 or B1.177 (SARS-CoV-2 lineages). A.27 ($n = 6$; 10.3%), B.1.1.7 ($n = 6$; 10.3%), B.1.1 ($n = 4$; 6.9%) and B.1.428.2 ($n = 3$; 5.2%) were the least frequent lineages. B.4, B.1.1.189 and B.1.540 were identified in tow patient each one ($n = 2$; 3.4%). B.55, B.1.9, B.1.177.6, B.1.333, B.1.356 and B.1.597 lineages were identified in one patient each (Fig. 1). The geographical distribution of SARS-CoV-2 lineages in Tunisia showed that B1.160 and B1.177 lineages circulated in most governorates. The Alpha variant B.1.1.7 has been identified mainly in Kasserine but also in Tunis and Ariana. The governorates of Tunis, Ben Arous, and Nabeul showed a higher diversity of lineage than the other governorates. A local cluster of the A.27 lineage was detected in Zaghouan and Sousse (Supplementary data Fig. S1).

3.2. SARS-CoV-2 phylogenetic and clades genetic characterization

A total of 29,424 positions were found in the final dataset. According to the Nextclade nomenclature, the 58 Tunisian isolates were dispersed in 7 different SARS-CoV-2 subclades including 19A, 19B, 20A, 20B, 20C,

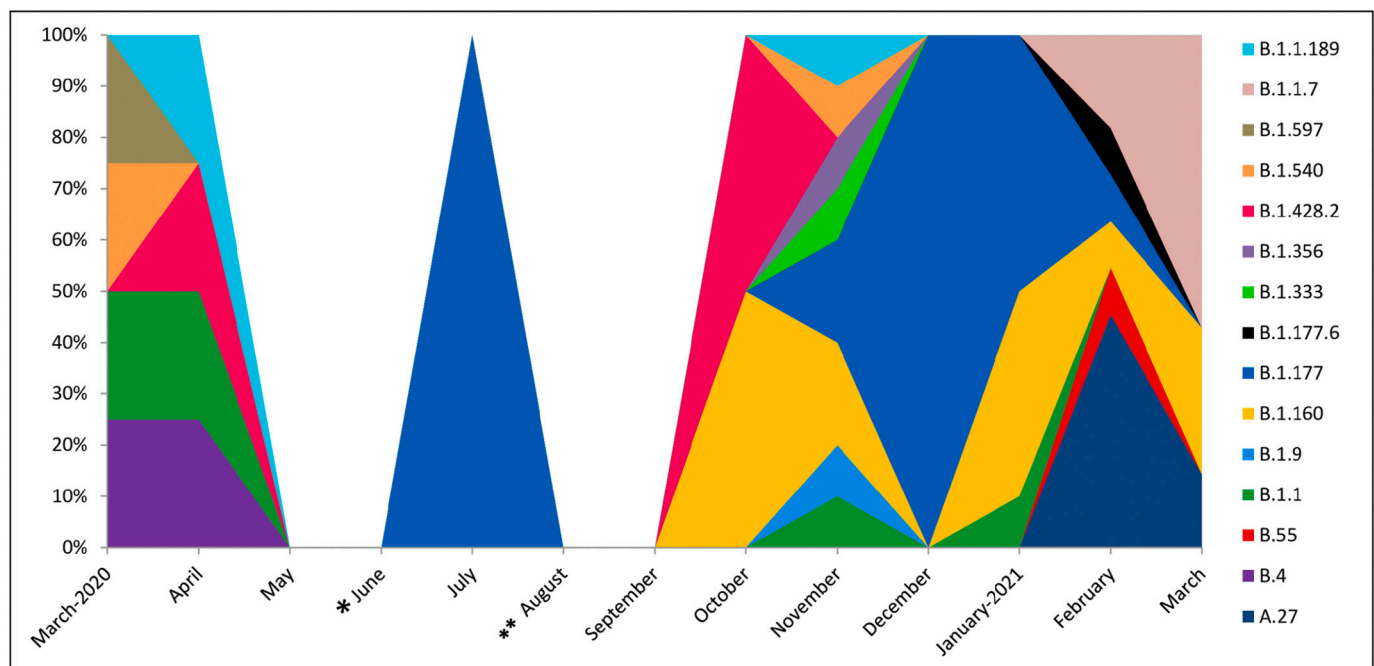


Fig. 1. Distribution of the 58 Tunisian SARS-CoV-2 genetic lineages in Tunisia during the first year of the pandemic: March 2020 to March 2021. *No COVID-19 cases were detected in Tunisia during June 2020, ** According to our sampling strategy, no COVID-19 cases were included during August 2020.

20 E and 20I (Fig. 2). The first two sub-clusters 19A and 19B were clustered together from a single node C1. The second node C2 contained sequences belonging mainly to the 2020 clades (20A, 20B, 20C, 20E and 20I). The similarity score between C1 and C2 ranged between 99.81% and 99.97% relative to the reference sequence. The two main nodes shared 4 non-synonymous mutations, one in ORF1ab gene (L3606F), and three in Spike protein (L18F, N501Y and K1191N) (Supplementary Data Tables S3–S4).

The 19A subclade gathered the Wuhan reference sequence with sequences “6736” and “7899” collected in the first pandemic period (March and April 2020). These sequences showed an identity score of 97.97% when compared to the NC_045512. The difference between these 2 and the reference sequence was in 8 locations affecting the nonstructural proteins NSP2 ($n = 2$), NSP4 ($n = 2$), and NSP6 ($n = 1$), and the structural N-Protein ($n = 3$). Among these, 5 caused changes in the protein sequences of ORF1a (V378I, G3072C, and L3606F) and N

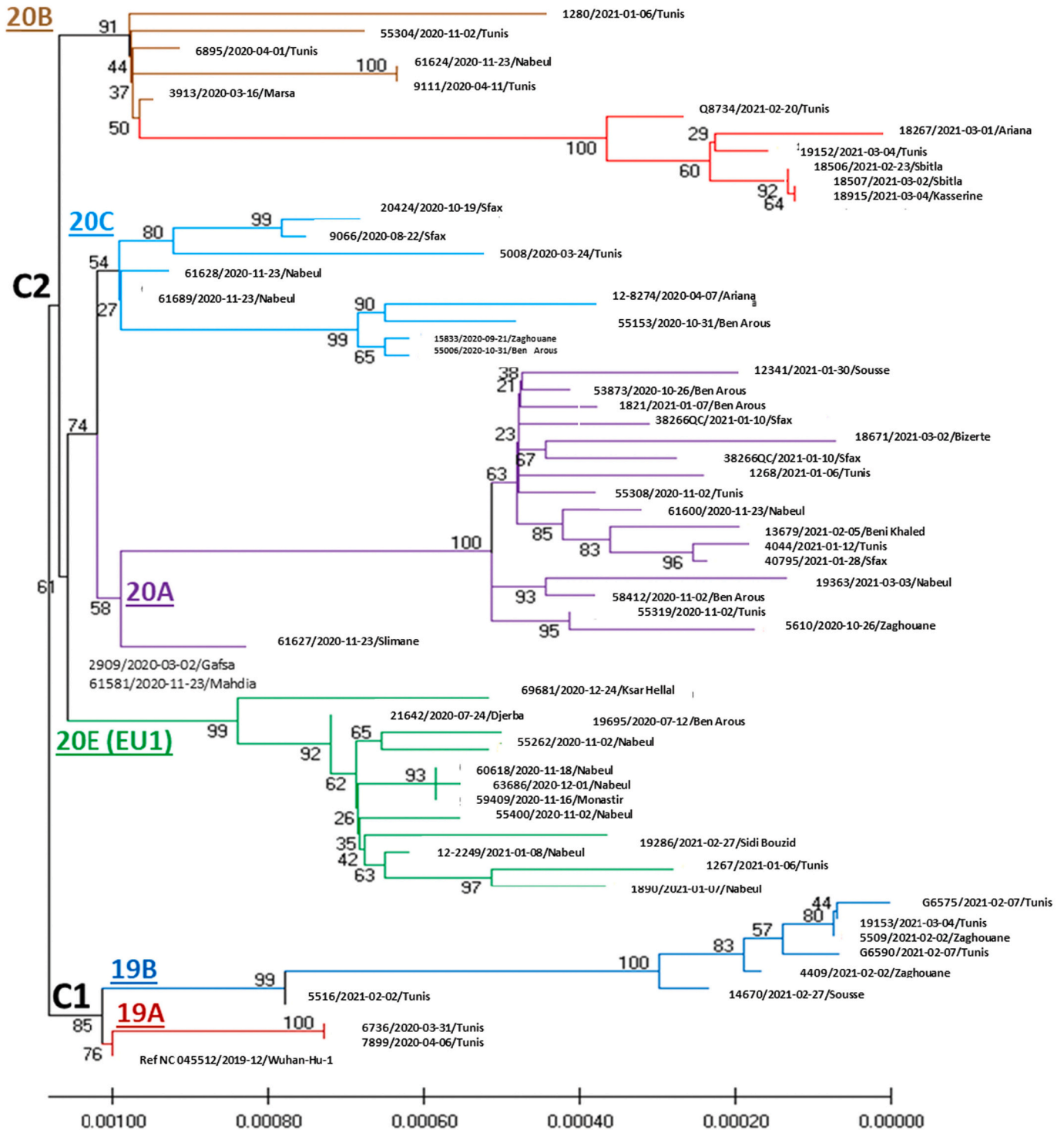


Fig. 2. Phylogenetic analysis of the 58 Tunisian SARS-CoV-2 genome sequences. Phylogenetic analysis of 58 Tunisian SARS-CoV-2 sequences, compared with SARS-CoV-2 reference sequence of Wuhan*: NC_045512, inferred by Neighbor-Joining method. Branches are colored according to the Nexclade Clade Nomenclature. The evolutionary distances were computed using the Maximum Likelihood method.

(M1I and S188P) (Table 1, Supplementary Data Tables S3–S4). The subclade 19B, groups strains isolated between February and March 2021 and has a similarity score ranging from 99.81% to 99.87% to the Wuhan reference. The variability profile of 19B subclade sequences was also characterized by the emergence of mutation sets affecting ORF1ab ($n = 12$), S-Protein ($n = 13$), ORF3a ($n = 2$), ORF8 ($n = 5$) and N-Protein ($n = 2$). The 6 nucleotide deletions in ORF8 were responsible for 2 amino acid deletions “D119” and “F120”.

The second major node C2 revealed that the Tunisian SARS-CoV-2 sequences were different from the reference sequence and were split into 5 clades 20B, 20I, 20C, 20A and 20E (Fig. 2), all sharing the spike mutations D614G and NSP12-RdRp mutation P314L (Table 1, Supplementary Data Tables S3–S4). 20A and 20E represented the 2 main subclades (in purple and green) (Fig. 2) bringing together sequences of SARS-CoV-2 viruses isolated at different times and locations in Tunisia. Two mutations, T223I and H1101Y/L5F in ORF3a and S-Protein, respectively, were shared between these subclades. Moreover, in February 2021, the clade 20I (Alpha, V1) emerged (red cluster). This

clade shares mutations in the N-Protein (R203K and G204R) with clade 20B. A unique mutation in the clades that constitute their genetic signature was also observed (Table 1, Supplementary Data Tables S3–S4). Mainly in 20I (Alpha, V1) with 18, 19B with 14, 19A with 7, 20E with 6, 20A with 4, 20C with 1 and 20B with none.

3.3. SARS-CoV-2 genomic characterization

The multiple sequence alignment of the fifty-eight Tunisian sequences according to the Wuhan-Hu-1 reference sequence (NC_045512) revealed a total of 335 mutation events including 325 single nucleotide polymorphisms (SNPs) and 10 deletions. Among the 325 SNPs, 239 transitions and 86 transversions were observed. These variations represented 134 synonymous and 191 non-synonymous changes (Table 1, Supplementary Data Table S4). Among all amino acid changes, 62 were found in structural proteins where 38 were observed in the spike (S) glycoprotein; 22 in the nucleocapsid (N) and only 2 in the membrane protein (M). No changes were seen in the envelope protein (E). Ninety-

Table 1
Non-synonymous mutations among SARS-CoV-2 clades from 58 Tunisian samples.

| | | 20B (N = 6) | 20I (Alpha) (N = 6) | 20C (N = 6) | 20 A (N = 17) | 20 E (N = 14) | 19 B (N = 7) | 19 A (N = 2) |
|-----------|--|--|---|--|---|--|--|---------------------|
| ORF1a | NSP1 | | | | E93K ¹ , R124C ¹ , G192D ¹ | | | |
| | NSP2 | G392C ¹ , E489D ¹ | L730F ⁴ | T265I ⁶ , S318L ¹ , T346I ³ , H388Y ³ , A482V ¹ | K292R ¹ , E342G ¹ , H1141Y ¹ , M1312I ¹ | D194N ¹ | P286L ⁶ | V378I ² |
| | NSP3 | A591V ¹ | T1001I ⁶ | P1596L ³ , T1908I ¹ , T2154I ¹ , S2535L ¹ , A2690V ³ | P1596L ¹ , K1895N ¹ , L2688F ¹ , M3087I ¹⁴ , L3201I ¹⁰ | V559M ¹ , K1247N ¹² , S1515F ¹ , P1659S ¹ , P1803S ² | | |
| | NSP4 | P1158S ¹ | A1708D ⁶ | T3058I ¹ | T3284I ¹ , K3353R ¹ | | D2980G ⁶ , T3082I ⁴ | G3072C ² |
| | NSP5 | I1232V ¹ , P3359S ² | I2230T ⁶ | | S3384L ¹ , L3711F ¹ , P4223L ¹ | P2018S ¹ | S3386F ¹ | |
| | NSP6 | C2210F ² , T3716A ¹ | S367S ⁻⁶ , G367G ⁻⁶ , F3677 ⁻⁶ | | | A2345V ¹ , A3209V ¹ , A3497V ¹ , L3606F ² | N3651S ⁶ | L3606F ² |
| | NSP7 | S2500F ² , S3884L ¹ | | | | | | |
| | NSP8 | | | | | A3623S ¹² | | |
| | NSP9 | | | | | | P4197S ¹ | |
| | NSP10 | | | T4304I ¹ | | | | |
| ORF1b | NSP12 | P314L ⁶ | T132I ¹ , P314L ⁶ | P314L ⁶ | A176S ¹⁴ , P314L ¹⁷ , T730I ¹ , V767L ¹⁴ , S904L ¹ | D275Y ² , P314L ¹⁴ | | |
| | NSP13 | | K1383R ² | M1156I ¹ , M1499I ¹ | P976L ¹ , K1141R ¹⁴ , E1184D ¹⁴ , M1352I ² , S1408L ¹ | P1001S ¹ , Y1229C ¹ | P1000L ⁶ | |
| | NSP14 | T1545A ¹ | E1871G ² | | T1540I ¹ , T1555I ¹ , R1737L ¹ , T1747N ¹ , V1840F ¹ , T2040I ¹ | | | |
| | NSP15 | | | M2269I ¹ | D2179Y ¹ , Q2247H ² , E2253G ¹ , D2263N ¹ | P2313S ¹ | | |
| | NSP16 | Q2635H ² | | | | A2559S ¹ | | |
| | S-Protein | D111N ¹ , D614G ⁶ , A684V ² , I770V ¹ , A892S ¹ | V6A ² , H69 ⁵ , V70 ⁵ , D138H ¹ , Y144 ⁶ , N501Y ⁶ , A570D ⁶ , D614G ⁶ , P681H ⁶ , T716I ⁶ , S982A ⁶ , D1118H ⁶ | D614G ⁶ , E780Q ¹ | L5F ¹ , V120I ¹ , L176F ¹ , I233V ² , G261S ¹ , S477N ¹⁴ , D614G ¹⁷ , D627E ¹ , S640F ¹ , D936Y ¹ , A1020S ¹ , H1101Y ¹ | A2559S ¹ , L5F ² , L18F ¹ , Q23H ¹ , A222V ¹⁴ , T572I ¹ , D614G ¹⁴ , G932C ¹ , H1101Y ¹ , K1191N ³ | L18F ⁶ , V227A ⁴ , L452R ⁶ , N501Y ⁶ , A653V ⁶ , H655Y ⁶ , Q677H ¹ , D796Y ⁶ , K1191N ⁵ , G1219V ⁶ | |
| ORF3a | F230V ¹ | L140F ¹ , G174D ³ | Q57H ⁶ , A99V ¹ , G224C ¹ | Q57H ¹⁵ , V97I ¹ , T223I ¹ | A39T ¹ , D155Y ¹ , L101F ¹ , W131C ¹ , T223I ² , V70F ¹ , I10L ¹ | V50A ¹ | | |
| M-Protein | | | | | | | | |
| ORF7a | F6S ¹ | | T14I ¹ | H125Y ¹ , T120I ¹ | | | | |
| ORF8 | S21N ² | Q27* ⁶ , R52I ⁶ , Y73C ⁵ | | H28Y ¹ , A51S ¹ , Q72L ¹ , C83F ¹ | V5I ¹ , E64* ¹ | D119* ⁶ , F120* ⁶ , A65S ⁴ , L84S ⁶ | | |
| ORF9b | | | | | S6I ¹ , Q18H ¹ | | | |
| N-Protein | R203K ⁶ , G204L ¹ , G204R ⁵ , S5G321F ² , T325I ¹ | D3L ⁵ , R203K ⁶ , G204R ⁶ , S235F ⁶ | S186Y ¹ , T205I ³ , D348H ¹ | K80E ³ , G19R ¹ , Q83R ³ , M234I ¹⁴ , A376T ¹⁴ , Q384H ² | Q9H ¹ , D22Y ¹⁴ , A220V ¹ , G236C ¹ , Q418L ¹ | S202N ⁶ | M1I ² , S188P ² | |

Superscript number: number of isolates that harboured mutation; - : deletion; *: stop codon; E: envelope protein; M: membrane glycoprotein; N: nucleocapsid phosphoprotein; ORF: open reading frame; S: spike glycoprotein.

six additional amino acid changes were identified in non-structural proteins (NSPs 1–16 in ORF1ab), and 33 in accessory protein genes such as ORF3a ($n = 13$), ORF7a ($n = 4$), ORF8 ($n = 13$) and ORF9b ($n = 3$) (Table 1, Supplementary Data Tables S3–S4). Deletions were observed at 10 sites and were identified in the genomic sequences of NSP3, NSP 6, protein S, protein E, ORF7a, ORF8 and protein N. Seven genomic deletions caused consequently the deletion of amino acids and six others caused frameshifts (Table 2).

3.4. Hyper-variable genomic hotspots

Interestingly, among the 191 non synonymous mutations, 17 were found as hotspots with more than 20% of mutation frequency in SARS-CoV-2 genomes derived from Tunisian patients. Eight out of them were found in ORF1ab, four in the spike glycoprotein (S), four in the nucleocapsid and one in accessory protein of ORF3a. In addition, some variants which presented the P314L in ORF1b-non-structural protein RNA-dependent RNA polymerase (RdRp), the D614G in Spike and the Q57H in ORF3a mutations, have a frequency that exceeded 30% in Tunisian sequences (Fig. 3).

4. Discussion

Face to the unusual SARS-CoV-2 pandemic, our country has increased its genomic capacity to track the rapid evolution of the virus. A national strategy which includes federated multi-disciplinary research projects (PRF) has been implemented during the early pandemic stage. The participation of the National Observatory of New and Emerging Diseases was to define sampling strategies during the different waves of pandemic and to capture strains with particular priority for sequencing. The first strain introduced in Tunisia was completely sequenced in the framework of this national project. The implemented national strategy also allowed the detection of different VOCs as soon as they were introduced in the country (<https://www.gisaid.org/>).

Our data revealed that the Tunisian SARS-CoV-2 pandemic started with multiple introduction events. Indeed, among the 14 different lineages generated in this study six were identified during the early phase including B.1, B.1.1, B.1.1.189, B.1.428.2, B.1.597 and B4. During this period the number of positive cases in Tunisia had reached 75 with 3 deaths. Preventive strategies were implemented, particularly, lockdown during 75 days and enhancement of contact tracing around all positive cases (NONED, 28 March, 2020). Since then, (Jun, 2020), the country recorded zero new cases for 15 consecutive days. These findings support the efficacy of the implemented strategies to contain the spread of SARS-CoV-2 in the country. Following the opening of the frontiers on

Jun 27, 2020, the epidemic situation in Tunisia changed with diverse new introduction of SARS-CoV-2 lineages. As the majority of European countries B.1.160, B.1.177 and B.1.1.7 were the predominant lineages during the second and the third waves (Hodcroft et al., 2021).

Our data revealed a clear difference between the North of the country (Tunis, Nabeul, Ben Arous and Ariana) where diverse lineages circulated and the South (Gafsa, Djerba, Sidi Bouzid) where only one lineage was exclusively present. Of note, Tunis, the most important urban area of Tunisia, is the only governorate where the majority of lineages (9/14) circulated. Given the central situation of Tunis and its intense economical the simultaneous circulation of the different lineages in this area would seem plausible.

Our results identified 17 hyper-variable genomic hotspots with frequencies above 20%. These hotspots and other mutations constitute genetic signatures that have allowed us to conduct the classification of the sequenced SARS-CoV-2 viruses. Among the 11 described clades (Rambaut et al., 2020), only 7 were identified in our sequence analysis. Sequence analysis allowed identifying 239 transitions consisting in an interchange between purine (A, G) or pyrimidine (C, T) bases and 86 transversions consisting in an interchange of purine into pyrimidine bases involving structural change. Many SARS-CoV-2 studies observed the predominance of transition type particularly when C is replaced by T (Koyama et al., 2020; Wang et al., 2020). The occurrence of such change is likely to reflect the virus adaptation process in its hosts (Matyášek and Kovařík, 2020). Among the 191 mutations/aa changes, the 2 most frequent mutations, D614G in S protein and P314L in Nsp12 appeared simultaneously and are often associated with increased viral infectivity (Korber et al., 2020). They were shared by all classified clades except 19A and 19B clades explaining the likely occurrence during the transition between the first cluster reported in Wuhan and the subsequent clusters that spread globally. Moreover, the high frequency of these mutations shows that they probably improved the fitness of the virus (Plante et al., 2021) by facilitating its functional cooperation and maintaining the stability of the virus, thus increasing the transmission rate and improving the adaptability of the virus (Laha et al., 2020). The third most frequent mutation Q57H (36.2%) in ORF3a causes the appearance of an early stop codon resulting in a truncated form of the accessory protein ORF3b (13 amino acid instead of 57). The ORF3b-Q57H-SARS-CoV-2 variant was found to be responsible for the fourth epidemic wave of COVID-19 in Hong Kong (Chu et al., 2021). Indeed, ORF3b protein is an interferon antagonist, and its functional failure might decrease its virulence and increase its transmissibility (Lam et al., 2020; Chu et al., 2021) demonstrated that variants carrying this mutation could evade induction of cytokine, chemokine, and interferon-stimulated gene expression in primary human respiratory

Table 2
Deletion characteristics: 58 Tunisian SARS-CoV-2 whole genome sequences.

| Deletions | Number of nucleotides | Position (bp) | AA change | Frameshift | Corresponding Protein | Clades | Number of affected sequences | Sequences reference |
|-------------|-----------------------|---------------|-----------|------------|-----------------------|-----------------|------------------------------|--------------------------------------|
| Deletion 1 | 7 | 6833 | I-K | Yes | Nsp3 | 20 E | 1 | 55,400 |
| Deletion 2 | 9 | 11,288 | S-G-F | No | Nsp6 | 20I (Alpha. V1) | 6 | Q8734/18267/19152/18506/18507/18915 |
| Deletion 3 | 6 | 21,765 | H-H | No | Protein S | 20I (Alpha. V1) | 6 | Q8734/18267/19152/18506/18507/18915 |
| Deletion 4 | 3 | 21,992 | Y | No | Protein S | 20I (Alpha. V1) | 6 | Q8734/18267/19152/18506/18507/18915 |
| Deletion 5 | 4 | 26,158 | V | Yes | Protein E | 20B | 1 | 55,304 |
| Deletion 6 | 8 | 26,161 | N-P | Yes | Protein E | 19B | 6 | G6590/19153/G6575bis/5509/4409/14670 |
| Deletion 7 | 1 | 27,293 | – | Yes | Orf7a | 20A | 1 | 55,319 |
| Deletion 8 | 1 | 27,388 | – | Yes | Orf7a | 19B | 6 | G6590/19153/G6575bis/5509/4409/14670 |
| Deletion 9 | 6 | 28,248 | D-F | No | Orf8 | 19B | 6 | G6590/19153/G6575bis/5509/4409/14670 |
| Deletion 10 | 1 | 28,271 | – | Yes | Protein N | 20I (Alpha. V1) | 5 | 18,267/19152/18506/18507/18915 |

Bp: Base pairs; AA: Amino Acids; Nsp: Non structural protein; Protein S: Spike glycoprotein; Protein E: Envelopeprotein; Protein N: Nucleocapsid protein.

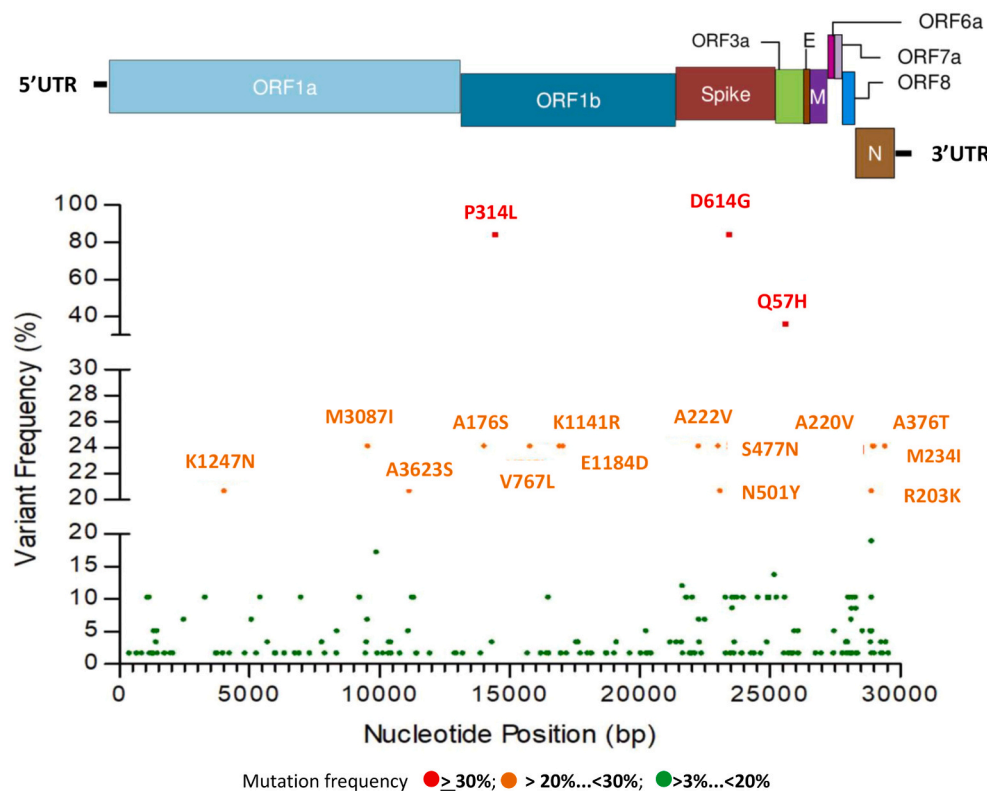


Fig. 3. Genomic variation frequency of Tunisian SARS-CoV-2 sequences ($n = 58$). Genomic variants were identified by referring to the first diagnosed NC_045512 Wuhan variant using MEGA X (Kumar et al., 2018). The locations and the mutations frequencies of the variants were plotted along genomic sequence of NC_045512. The open reading frames (ORFs) of SARS-CoV-2 were shown as rectangles that were aligned with nucleotide positions of the coronavirus. The frequency of each mutation in the population is presented by color coded circles. Abbreviations: ORF: Open Reading Frame; E: Envelope; M: Membrane protein; N: Nucleocapsid protein.

cells. This mutation is found in an increasing number of sequences and coincided with the appearance of the D614G mutation. A hypothesis was raised on the link between the two mutations acting simultaneously. Other less frequent hotspots were often associated with specific clades like R203K mutation in M protein found in clades 20B and 20I which was almost always associated with G204R and have been previously correlated with enhanced virulence (Wu et al., 2021). A222V mutation in S protein found in area defined as possible immune cells epitope were described elsewhere in 20 E clade (Bao-zhong Zhang et al., 2020).

Moreover, six relatively recent sequences (collected in February and March 2021) corresponded to 19B clade which appeared in early 2020, was expected to disappear over time (Murall et al., 2021), but which frequency increased worldwide probably due to convergent mutations affecting S protein including D614G (Volz et al., 2021). In our sequences we did not observe any D614G mutation in this clade, but rather a combination of L18F, N501Y, L452R, H655Y, D796Y and G1219V mutations in S protein. These later, (L18F, N501Y, L452R and H655Y) were previously described in France (Fourati et al., 2021) and were suggested to improve the interaction of S protein with ACE2 viral receptor (Weisblum et al., 2020) and conducting to increased resistance to neutralizing antibodies (Choi et al., 2020; Baum et al., 2020). Moreover, 3 deletion sites causing 8 nucleotides missing in protein E, a frameshift in ORF7a and 6 nucleotides missing in ORF8 were also found in all 19 B sequences. Protein E is involved in several aspects such as assembly, budding, envelope formation and pathogenesis (Schoeman and Fielding, 2019; Chai et al., 2021). Accessories proteins, ORF7a and ORF8, interfere with host protein and mediate immune response. Their function alteration may influence the virulence (Tse et al., 2021 and Zhang et al., 2020). Any such event resulting in the relative re-emergence of a clade should be closely monitored in the future. Notably, 3 deletions causing frameshift in NSP3, protein E and ORF7a were affecting only one sequence and will be verified with Sanger sequencing. Likewise, it would be interesting to study *in vitro* and *in vivo* the functional consequences of those deletions on the virus if confirmed. The rate of mutations causes genome dynamics, flexibility, plasticity and variability, all

leading to virus evolution, and the spread of polymorphic variants. Most mutations that were responsible for variants classification occurred in S protein, but several other emerging mutations were observed in both coding and non-coding sequences of other genes especially NSP12, NSP9 and N proteins (Khailany et al., 2020). In this study, we identified and characterized the Alpha variant (clade 20I) or B.1.1.7 by its characteristic mutation and deletion affecting the S protein (69 del, 70 del, 144 del, N501Y, A570D, D614G, P681H, T716I, S982A and D1118H) (Meng et al., 2021). The deletion 69–70 probably alters the N-terminal domain loop conformation in the S protein and could be associated with increased infectivity (Kemp et al., 2020). In addition, one previously non described one nucleotide deletion in N protein was also found in five of the six alpha variant sequences. This deletion might involve a frameshift in the sequence. According to the Center for Disease Control and Prevention, alpha variant is categorized among the variants of concern (VOC) which contains variants presenting diagnostic interferences and increased resistance to therapy or vaccination in addition to evidence of increased transmissibility and disease severity. This variant first appeared in Tunisia on 2 March 2021 and its highly accelerated transmission (Davies et al., 2021) allowed it to be in the first rank during the third wave of COVID-19 pandemic in spring 2021 Tunisia. It affected many cases and put health facilities under strain. It is important to notice that the mutation K1191N observed in the alpha variant in the USA (Washington et al., 2021) was also identified in our 20E and 19B clades.

As seen, the appearance of some mutations and/or variants can be a major event in the epidemic evolution and the spread of genetically polymorphic variants. Continuing to sequence new variants as soon as they appear could therefore be useful, in monitoring and managing the epidemy, as well as treatment and vaccine development.

Data availability

Genome sequences generated in this study were deposited in the GISAID (<https://www.gisaid.org>) and GenBank (<https://www.ncbi.nlm.nih.gov>) databases. Accession IDs are available in Supplementary

Table Data S2.

Funding

This work was supported by the Tunisian Ministry of High Education and Research: Federated Research Projects: PRFCOVID-GP3.

Sequence data

Sequence data from this article have been deposited in GISAID (<https://www.gisaid.org>) and GenBank (<https://www.ncbi.nlm.nih.gov>) databases.

CRedit authorship contribution statement

Mouna Ben Sassi: Investigation, Resources, Data curation, Writing – original draft, Validation. **Sana Ferjani:** Conceptualization, Investigation, Resources, Formal analysis, Data curation, Visualization, Writing – original draft, Validation. **Imen Mkada:** Conceptualization, Resources, Data curation, Formal analysis, Writing – original draft, Visualization, Validation. **Marwa Arbi:** Conceptualization, Resources, Data curation, Formal analysis, Writing – original draft, Visualization, Validation. **Mouna Safer:** Methodology, Formal analysis, Validation. **Awatef Elmoussi:** Investigation, Resources, Validation. **Salma Abid:** Investigation, Resources, Validation. **Oussema Souiai:** Conceptualization, Formal analysis, Resources, Data curation, Writing – review & editing, Validation. **Alya Gharbi:** Resources, Validation. **Asma Tejouri:** Validation. **Emna Gaies:** Validation. **Hanene Eljabri:** Validation. **Samia Ayed:** Resources, Validation. **Aicha Hechaichi:** Validation. **Riadh Daghfous:** Validation. **Riadh Gouider:** Project administration, Supervision, Validation. **Jalila Ben Khelil:** Supervision, Validation. **Maher Kharrat:** Supervision, Validation. **Imen Kacem:** Resources, Validation. **Nissaf Ben Alya:** Methodology, Validation. **Alia Benkahla:** Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – review & editing, Supervision, Validation. **Sameh Trabelsi:** Supervision, Validation. **Ilhem Boutiba-Ben Boubaker:** Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Validation.

Declaration of Competing Interest

The authors declare that there is no competing interest.

Acknowledgements

Authors gratefully acknowledge the World Health Organization and the European project STAMINA (<https://cordis.europa.eu/project/id/883441/fr>) for the financial support of young researchers: Sana Ferjani and MarwaArbi.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2022.105300>.

References

- Abid, S., Ferjani, S., El Moussi, A., Ferjani, A., Nasr, M., Landolsi, I., Saidi, K., Gharbi, H., Letaief, H., Hechaichi, A., Safer, M., Ben Alaya, N., Boubaker, I.B., 2020. Assessment of sample pooling for SARS-CoV-2 molecular testing for screening of asymptomatic persons in Tunisia. *Diagn. Microbiol. Infect. Dis.* 98 (3), 115125 <https://doi.org/10.1016/j.diagmicrobio.2020.115125>.
- Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., Wei, Y., Atwal, G.S., Murphy, A.J., Stahl, N., Yancopoulos, G.D., Kyrtasous, C.A., 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* (New York, N.Y.) 369 (6506), 1014–1018. <https://doi.org/10.1126/science.abd0831>.

- Bian, L., Gao, F., Zhang, J., He, Q., Mao, Q., Xu, M., Liang, Z., 2021. Effects of SARS-CoV-2 variants on vaccine efficacy and response strategies. *Expert Rev. Vaccines* 20 (4), 365–373. <https://doi.org/10.1080/14760584.2021.1903879>.
- Chai, J., Cai, Y., Pang, C., Wang, L., McSweeney, S., Shanklin, J., Liu, Q., 2021. Structural basis for SARS-CoV-2 envelope protein recognition of human cell junction protein PALS1. *Nat. Commun.* 12 (1), 3433. <https://doi.org/10.1038/s41467-021-23533-x>.
- Chakroun, H., Ben Lasfar, N., Fall, S., Maha, A., El Moussi, A., Abid, S., Rouis, S., Bellazreg, F., Abassi-Bakir, D., Bouaffif Ben Alaya, N., Boutiba Ben Boubaker, I., Hachfi, W., Letaief, A., 2020. First case of imported and confirmed COVID-19 in Tunisia. *La Tunisie medicale* 98 (4), 258–260.
- Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., Solomon, I.H., Kuo, H.H., Boucau, J., Bowman, K., Adhikari, U.D., Winkler, M.L., Mueller, A.A., Hsu, T.Y., Desjardins, M., Baden, L.R., Chan, B.T., Walker, B.D., Lichterfeld, M., Brigl, M., Li, J.Z., 2020. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Engl. J. Med.* 383 (23), 2291–2293. <https://doi.org/10.1056/NEJMc2031364>.
- Chu, D., Hui, K., Gu, H., Ko, R., Krishnan, P., Ng, D., Liu, G., Wan, C., Cheung, M.C., Ng, K.C., Nicholls, J.M., Tsang, D., Peiris, M., Chan, M., Poon, L., 2021. Introduction of ORF3a-Q57H SARS-CoV-2 Variant Causing Fourth Epidemic Wave of COVID-19, Hong Kong, China. *Emerg. Infect. Dis.* 27 (5), 1492–1495. <https://doi.org/10.3201/eid2705.210015>.
- Chunyang, L., Fang, J., Liang, W., Liping, W., Jungui, H., Mingjia, D., Yan, L., Xiucheng, P., Juanjuan, F., Li, L., Guangde, Y., Jianye, Y., Xuebing, Y., Bing, G., 2020. Asymptomatic and Human-to-Human Transmission of SARS-CoV-2 in a 2-Family Cluster, Xuzhou, China. *Emerg. Infect. Dis.* 26 (7), 1626–1628. <https://doi.org/10.3201/eid2607.200718>.
- Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waite, W., Wong, K., van Zandvoort, K., Silverman, J.D., CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz, K., Keogh, R., Edmunds, W.J., 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* (New York, N.Y.) 372 (6538), eabg3055. <https://doi.org/10.1126/science.abg3055>.
- Fourati, S., Decousser, J.W., Khouider, S., N'Debi, M., Demontant, V., Trawinski, E., Gourgeon, A., Gangloff, C., Destras, G., Bal, A., Josset, L., Soulier, A., Costa, Y., Gricourt, G., Lina, B., Lepeule, R., Pawlowsky, J.M., Rodriguez, C., 2021. Novel SARS-CoV-2 variant derived from clade 19B, France. *Emerg. Infect. Dis.* 27 (5), 1540–1543. <https://doi.org/10.3201/eid2705.210324>.
- Hodcroft, E.B., Zuber, M., Nadeau, S., Vaughan, T.G., Crawford, K., Althaus, C.L., Reichmuth, M.L., Bowen, J.E., Walls, A.C., Corti, D., Bloom, J.D., Veessler, D., Mateo, D., Hernandez, A., Comas, I., González Candelas, F., SeqCOVID-SPAIN consortium, Stadler, T., Neher, R.A., 2021. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv*. <https://doi.org/10.1101/2020.10.25.20219063> (the preprint server for health sciences, 2020.10.25.20219063).
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* (Oxford, England) 28 (12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Kemp, S.A., Datt, R.P., Collier, D.A., Ferreira, I.A.T.M., Carabelli, A., Harvey, W., Robertson, D.L., Gupta, R.K., 2020. Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion Δ H69/ Δ V70. *Cell Rep.* <https://doi.org/10.1016/j.celrep.2021.109292>.
- Khailany, R.A., Safdar, M., Ozaslan, M., 2020. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* 19, 100682 <https://doi.org/10.1016/j.genrep.2020.100682>.
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Sheffield COVID-19 Genomics Group, McDanal, C., Perez, L.G., Tang, H., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182 (4), 812–827.e19. <https://doi.org/10.1016/j.cell.2020.06.043>.
- Koyama, T., Platt, D., Parida, L., 2020. Variant analysis of SARS-CoV-2 genomes. *Bull. World Health Organ.* 98 (7), 495–504. <https://doi.org/10.2471/BLT.20.253591>.
- Kumar, S., Stecher, G., Li, M., Niyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Laha, S., Chakraborty, J., Das, S., Manna, S.K., Biswas, S., Chatterjee, R., 2020. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* 85, 104445 <https://doi.org/10.1016/j.meegid.2020.104445>.
- Lam, J.Y., Yuen, C.K., Ip, J.D., Wong, W.M., To, K.K., Yuen, K.Y., Kok, K.H., 2020. Loss of orf3b in the circulating SARS-CoV-2 strains. *Emerging Microbes Infections* 9 (1), 2685–2696. <https://doi.org/10.1080/22221751.2020.1852892>.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* (Oxford, England) 23 (21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
- Matyášek, R., Kovářik, A., 2020. Mutation patterns of human SARS-CoV-2 and bat RaTG13 coronavirus genomes are strongly biased towards C>U transitions, indicating rapid evolution in their hosts. *Genes* 11 (7), 761. <https://doi.org/10.3390/genes11070761>.

- Meng, B., Kemp, S.A., Papa, G., Dahir, R., Ferreira, I., Marelli, S., Harvey, W.T., Lytras, S., Mohamed, A., Gallo, G., Thakur, N., Collier, D.A., Mlcochova, P., COVID-19 Genomics UK (COG-UK) Consortium, Duncan, L.M., Carabelli, A.M., Kenyon, J.C., Lever, A.M., De Marco, A., Saliba, C., Gupta, R.K., 2021. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep.* 35 (13), 109292. <https://doi.org/10.1016/j.celrep.2021.109292>.
- Murall, C.L., Fournier, E., Galvez, J.H., N'Guessan, A., Reiling, S.J., Quirion, P.O., Sana, N., Anne, M.R., Shu, H.C., Paul, S., Mathieu, B., David, B., Romain, G., Pierre, L., Janick, C., Patrick, W., Réjean, D., Hugues, C., Mark, L., Michel, R., Guillaume, B., Jiannis, R., Jesse, Shapiro, Moreira, S., 2021. A small number of early introductions seeded widespread transmission of SARS-CoV-2 in Québec, Canada. medRxiv. <https://doi.org/10.1101/2021.03.20.21253835>.
- National Observatory of New and Emerging Diseases Ministry of Health, Tunisia, 2020. COVID-19 en Tunisie Point de situation à la date du 28 mars 2020. Available from. <http://www.omne.tn/fr/images/COVID-19-dernier-bulletin-28032020.pdf>.
- Parums, V., 2021. Editorial: revised World Health Organization (WHO) terminology for variants of concern and variants of interest of SARS-CoV-2. *Med. Sci. Monit.* 27, e933622 <https://doi.org/10.12659/MSM.933622>.
- Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Shi, P.Y., 2021. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592 (7852), 116–121. <https://doi.org/10.1038/s41586-020-2895-3>.
- Rader, B., White, L.F., Burns, M.R., Chen, J., Brilliant, J., Cohen, J., Shaman, J., Brilliant, L., Kraemer, M., Hawkins, J.B., Scarpino, S.V., Astley, C.M., Brownstein, J. S., 2021. Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. *Lancet. Digital Health* 3 (3), e148–e157. [https://doi.org/10.1016/S2589-7500\(20\)30293-4](https://doi.org/10.1016/S2589-7500(20)30293-4).
- Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5 (11), 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>.
- Schoeman, D., Fielding, B.C., 2019. Coronavirus envelope protein: current knowledge. *Virology* 16 (1), 69. <https://doi.org/10.1186/s12985-019-1182-0>.
- Sneath, P.H., Sokal, R.R., 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, 1st edn. W. H. Freeman, San Francisco.
- Thielen, P.M., Wohl, S., Mehoke, T., Ramakrishnan, S., Kirsche, M., Falade-Nwulia, O., Trovão, N.S., Ernlund, A., Howser, C., Sadowski, N., Morris, C.P., Hopkins, M., Schwartz, M., Fan, Y., Gnizdowski, V., Lessler, J., Sauer, L., Schatz, M.C., Evans, J. D., Ray, S.C., Mostafa, H.H., 2021. Genomic diversity of SARS-CoV-2 during early introduction into the Baltimore-Washington metropolitan area. *JCI Insight* 6 (6), e144350. <https://doi.org/10.1172/jci.insight.144350>.
- Tse, H., Wong, S.C., Ip, K.F., Cheng, V.C., To, K.K., Lung, D.C., Choi, G.K., 2021. Genome sequences of three SARS-CoV-2 ORF7a deletion variants obtained from patients in Hong Kong. *Microbiol. Resource Announc.* 10 (15), e00251-21 <https://doi.org/10.1128/MRA.00251-21>.
- Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., Rey, S.M., Nicholls, S.M., Colquhoun, R. M., da Silva Filipe, A., Shepherd, J., Pascall, D.J., Shah, R., Jesudason, N., Li, K., Jarrett, R., Connor, T.R., 2021. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* 184 (1), 64–75.e11. <https://doi.org/10.1016/j.cell.2020.11.020>.
- Wang, R., Hozumi, Y., Zheng, Y.H., Yin, C., Wei, G.W., 2020. Host immune response driving SARS-CoV-2 evolution. *Viruses* 12 (10), 1095. <https://doi.org/10.3390/v12101095>.
- Washington, N.L., Gangavarapu, K., Zeller, M., Bolze, A., Cirulli, E.T., Schiabor Barrett, K.M., Larsen, B.B., Anderson, C., White, S., Cassens, T., Jacobs, S., Levan, G., Nguyen, J., Ramirez, J.M., Rivera-Garcia, C., Sandoval, E., Wang, X., Wong, D., Spencer, E., Robles-Sikisaka, R., Andersen, K.G., 2021. Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. medRxiv, 2021.02.06.21251159. <https://doi.org/10.1101/2021.02.06.21251159>.
- Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J.C., Muecksch, F., Rutkowska, M., Hoffmann, H.H., Michailidis, E., Gaebler, C., Agudelo, M., Cho, A., Wang, Z., Gazumyan, A., Cipolla, M., Luchsinger, L., Hillyer, C.D., Caskey, M., Robbiani, D.F., Bieniasz, P.D., 2020. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* 9, e61312. <https://doi.org/10.7554/eLife.61312>.
- World Health Organization (WHO). <https://covid19.who.int/region/emro/country/tn>, 2021 (Accessed 30th, July 2021).
- World Health Organization (WHO). <https://covid19.who.int/>, 2021 (Accessed 30th, July 2021).
- Wu, S., Tian, C., Liu, P., Guo, D., Zheng, W., Huang, X., Zhang, Y., Liu, L., 2021. Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. *J. Med. Virol.* 93 (4), 2132–2140. <https://doi.org/10.1002/jmv.26597>.
- Zhang, B.Z., Hu, Y.F., Chen, L.L., Yau, T., Tong, Y.G., Hu, J.C., Cai, J.P., Chan, K.H., Dou, Y., Deng, J., Wang, X.L., Hung, I.F., To, K.K., Yuen, K.Y., Huang, J.D., 2020. Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Res.* 30 (8), 702–704. <https://doi.org/10.1038/s41422-020-0366-x>.