Title :

Sequencing using a two-steps strategy reveals high genetic diversity in the S gene of SARS-CoV-2 after a high transmission period in Tunis, Tunisia.

Authors:

Wasfi Fares[1*$], Kais Ghedira[2*], Mariem Gdoura[1,4], Anissa Chouikha[1], Sondes Haddad-Boubaker[1], Marwa Khedhiri[1], Kaouthar Ayouni[1], Asma Lamari[1], Henda Touzi[1], Walid Hammemi[1], Zina Medeb[1], Amel Sadraoui[1], Nahed Hogga[1], Nissaf ben Alaya[3,5], Henda Triki[1,5]

1. Laboratory of Clinical Virology - Reasearch Laboratory "Viruses Vectors and Hosts" (LR20-IPT10) - Institut Pasteur, University of Tunis-El Manar, Tunis, Tunisia

2. Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), Institut Pasteur de Tunis (IPT), 13, Place Pasteur BP 74, Tunis 1002, University of Tunis-El Manar, Tunis, Tunisia.

3. National Observatory for New and Emerging Diseases, Ministry of Health, Tunis, Tunisia

4. Faculty of Pharmacy, University of Monastir, Tunisia

5. Faculty of Medicine, University of Tunis-El Manar, Tunis, Tunisia

* Both authors have equally contributed to this work

$ Correspondence should be addressed to this author

E-mail addresses: Wasfi.fares@pasteur.tn;

Abstract:

Recent efforts have reported numerous variants that influence SARS-CoV-2 viral characteristics including pathogenicity, transmission rate and ability of detection by molecular tests. Whole genome sequencing based on NGS technologies is the method of choice to identify all viral variants; however, the resources needed to use these techniques for a representative number of specimens remain limited in many low and middle income countries. To decrease sequencing cost, we developed a couple of primers allowing to generate partial sequences in the viral S gene allowing rapid detection of numerous variants of concern (VOCs) and variants of interest (VOIs); whole genome sequencing is then performed on a selection of viruses based on partial sequencing results. Two hundred and one nasopharyngeal specimens collected during the decreasing phase of a high transmission COVID-19 wave in Tunisia were analyzed. The results reveal high genetic variability within the sequenced fragment and allowed the detection of first introduction in the country of already known VOCs and VOIs as well as others variants that have interesting genomic mutations and need to be kept under surveillance.

Importance:

The method of choice for SARS-CoV-2 variants detection is whole genome sequencing using NGS technologies. Resources for this technology remain limited in many low and middle income countries where it is not possible to perform whole genome sequencing for representative number of SARS-CoV-2 positive cases. In the present work, we developed a novel strategy based on a first partial sanger screening in the S gene including key mutations of the already known VOCs and VOIs for rapid identification of these VOCs and VOIs and helps to better select specimens that need to be sequenced by NGS technologies. The second step consisting in whole genome sequencing allowed to have a holistic view of all variants within the selected viral strains and confirmed the initial classification of the strains based on partial S gene sequencing.

Key words: COVID-19, SARS-CoV-2, whole genome sequencing, VOCs, VOIs, protein Spike, Tunisia

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is the causative agent of human coronavirus disease 2019 (COVID-19), was identified in Wuhan-China in December 2019 (1, 2). The outbreak of the coronavirus disease (COVID-19) rapidly spread worldwide; it was officially declared as pandemic by the World Health Organization (WHO) on March 11, 2020 (3) and now represents a tremendous threat globally.

SARS-CoV-2 is a single-stranded positive RNA virus, a member of the Beta coronavirus genus that also contains SARS-CoV and MERS-CoV. The first sequence of the virus was published in January 2020 (4). The structural genome region, located in the 3' part of the genome, encodes four structural proteins: spike (S), envelope (E), membrane (M) and nucleocapsid (N) (5). The S protein forms a trimer on the surface of the virion, it mediates virus attachment to the ACE-2 receptor and its entry to the host cells (6). The S Protein is composed of two sub-units, S1 containing the receptor-binding domain (RBD) and S2 that mediates membrane fusion (7). The S protein determines SARS-CoV-2 infectivity and transmissibility and is also the major antigen inducing protective immune response (8). Since the beginning of the COVID-19 pandemic, the S protein has been undergoing several mutations and it is highly important to follow the emergence of these variants and their biological, epidemiological and clinical significance. Early in the pandemic, variants of SARS-CoV-2 containing a D to G substitution in the 614 amino-acid residue of the S protein (D614G) were reported. This substitution increased receptor binding avidity and D614G mutants became dominant in many geographic regions (9-11). In December 2020, the United Kingdom reported a variant of concern (VOC), referred as B.1.1.7, with enhanced transmissibility within the population (12, 13). This variant became predominant in the UK and spread to more than 100 countries in the world. In January 2021, two other VOCs, referred as B.1.351 and B1.1.28, also with high transmissibility, were reported in South Africa and Brazil, respectively (14-16). Later, many other variants, classified as Variants Under Investigation (VUIs) were reported throughout the world. In addition to the increased transmissibility, it is suggested that some mutations in these variants may affect the performance of some diagnostic real-time PCR tests and reduce susceptibility to vaccine-induced neutralizing antibodies (9, 10, 17-22). Global tracking of these newly identified VOCs and VUIs as well as any other evolving SARS-CoV-2 variant, by genomic surveillance and rapid sharing of viral genomic sequences, is highly recommended in order to limit their spread and control the pandemic.

Nowadays, several classifications of SARS-Co V-2 strains in lineages or clades were proposed. Indeed, two different lineages, A and B, were proposed by the Phylogenetic Assignment of Named Global Outbreak (PANGO) lineage nomenclature, while a classification in 11 different clades (19-A, 19-B, 20-A to 20-I) was proposed by the Nextstrain resources and another classification in 9 clades (S, L, O, V, G, GH, GR, GRY and GV) was proposed by GISAID.

In Tunisia, the first case of SARS-CoV-2 infection was reported on March 03, 2020 (23). The country experienced a first wave of the COVID disease and, through setting up drastic nation-wide multi-sectoral measures to avoid international introduction of the virus and its spread within the population, COVID-19 incidence decreased in May-June 2020 to reach zero cases per day from the 4th to the 11th of June 2020. The national strategy included early detection of imported cases, quarantining of new confirmed cases as well as suspected cases and strict travel restrictions. After the sharp decrease of the disease incidence; a relaxation in the application of these measures by the general population, combined with decreased restrictions in international transportation, led to the re-introduction of the virus again and the establishment of a local transmission. In late July, COVID-19 incidence started to increase again and the country experienced a second wave with highest incidence in January 2021, associated with a high local transmission within the population. Starting from February 2021, the disease incidence together with mortality rates decreased again.

The present work reports the genomic features of SARS-CoV-2 sequences detected in Tunisia during the late phase of the second wave of the pandemy and reveals the co-circulation of several variants, some of which are already known as VOCs, others have interesting genomic mutations and need to be kept under surveillance.

Material and Methods

Nasopharyngeal samples.

A total of 201 SARS-CoV-2 positive nasopharyngeal samples, collected from individuals living in the four districts of Tunis capital, were included in this study. Sample collection was performed from January to March 2021, during the decreasing phase of the second wave of COVID-19 outbreak in Tunisia (**Figure1**). The study population includes symptomatic patients presenting with mild COVID clinical forms or with severe forms as well as asymptomatic individuals sampled after a contact with confirmed cases. The study population included 91 males and 110 females, their age ranged from 5 to 98 years. The samples were collected by the health teams from the Ministry of Health, at home for asymptomatic individuals and those with non-severe clinical symptoms, or at the health facility level for hospitalized patients. Samples were transported, refrigerated and within 24 hours, to the Pasteur Institute of Tunis where they were immediately processed for SARS-CoV-2 detection by specific real time reverse transcription polymerase chain reaction (RT-PCR) according to WHO approved protocols (24, 25).

Ethical statement

This work was performed in the frame of COVID-19 diagnostic effort, and all samples used for analysis were anonymized. This study was approved by the Bio-Medical Ethics Committee of the Pasteur Institute of Tunis, Tunisia réf. 2020/14/I/LR16IPT/V1

Primer design.

Primers were designed using PrimerDesign-M online software, available through https://www.hiv.lanl.gov/content/sequence/PRIMER_DESIGN/primer_design.html (26, 27), based on an alignment of 13451 SARS-CoV-2 complete genome sequences. Several points were considered such as melting temperatures, G+C percentage, entropy, complexity and nucleotide composition, in order to perfectly align with the SARS-CoV-2 sequence. The selected primers sequences were as follows: IPT_FW: (22964-22987) 5'-ATTTCAACTGAAATCTATCAGGCC-3' and IPT_REV: (23666-23647) 5'-CTGCACCAAGTGACATAGTG-3'. Indicated positions correspond to the sequence of Wuhan reference strain (accession number: NC045512). The designed primers allow the amplification of a 703-nucleotide-long region in the S gene holding key mutations, that includes the E484K, N501Y, A570D, D614G and P681H, recently identified as specific of the main VOCs and VUIs of SARS-CoV-2.

PCR amplification and sequencing in the S gene.

A volume of 140µl of nasopharyngeal samples was used for viral RNA extraction with viral RNA Mini Kit (Qiagen, Hilden, Germany) to give a final elution volume of 60µl of total RNA. The presence of SARS-CoV-2 RNA was determined by conventional reverse transcription PCR using the SuperScript®III One-Step RT-PCR System with Platinum® Taq DNA Polymerase kit (Invitrogen) in a 25µl reaction volume containing 12.5µl of 2X buffer, 0.5µl Rnasin (Promega), 1µl of each reverse and forward primers (10µM), 1µl Enzyme mix and 5µl of total extracted RNA. Optimized cycling conditions was performed as follows: Reverse transcription with initial incubation at 50°C for 30min and 94°C for 2min followed by 35 cycles, repeating denaturation at 94°C for 15sec, annealing at 54°C for 45sec and elongation at 72°C for 30sec, and final elongation at 72°C for 10min. Amplification products are first visualized by electrophoresis in agarose gels and then purified by the ExoSAP-IT method using the Exonuclease-I and the Shrimp Alkaline Phosphatase (Invitrogen). The purified amplicons were sequenced using the Big Dye Terminators v3.1 kit (Applied Biosystems) and the forward and reverse PCR primers. The resulting consensus sequences were deduced by aligning the forward and the reverse sequence of each isolate, excluding primer binding regions and are 618 nucleotides-long (positions 22988 to 23605 according to the Wuhan reference strain NC045512). They were submitted to the NCBI database under accession number MZ150010 - MZ150210.

Whole genome sequencing.

The QIAseq SARS-CoV-2 Primer Panel paired with the QIAseq FX DNA Library construction kits (Qiagen GmbH, Germany) were used for enriching and sequencing the entire SARS-CoV-2 viral genome. Extracted RNA from nasopharyngeal swabs was first depleted of ribosomal RNA using RiboZero rRNA removal Kit (Illumina, USA). The residual RNA was then converted to double stranded cDNA using random priming. Following cDNA synthesis, QIAseq SARS-CoV-2 Primer Panel kit was used including high fidelity multiplex PCR reaction yielding 400bp amplicons covering the full viral genome. The multiplexed amplicon pools were then converted to sequencing libraries by enzymatic fragmentation with 250bp fragment size, end repair and ligation to adapters with the QIAseq FX DNA Library construction kits. Thereafter, the constructed DNA library was purified and adapter-dimers were removed with Agencourt AMPure XP beads. The libraries were sequenced using Nextseq (Illumina Inc, USA) to generate 2x150 bp paired-end sequencing reads.

Sequences' raw data have been processed using fastqc version 0.11.9 for quality control (**https://www.bioinformatics.babraham.ac.uk/projects/fastqc/**). Low quality reads and adapters have been filtered using trimmomatic version 0.39 (28) with a Phred quality score of 30 as threshold. Genome consensus sequences were assembled by mapping on the SARS-CoV-2 reference genome of GenBank accession number NC045512 (Wuhan-Hu-1 isolate) using Spades assembler version 3.15.0 (29), with thresholds of 80% for nucleotide sequence coverage and 90% for nucleotide similarity. The obtained SARS-CoV-2 new sequences were submitted to the GISAID database (https://www.gisaid.org) (30, 31) with the following accession numbers: EPI_ISL_2035560, EPI_ISL_2035563, EPI_ISL_2035720, EPI_ISL_2035734, EPI_ISL_2035752, EPI_ISL_2035753, EPI_ISL_2035940 to EPI_ISL_2035949, EPI_ISL_2035988 and EPI_ISL_2036077.

Phylogenetic analysis.

The obtained partial S gene sequences and selective whole genome sequences were aligned together with representative SARS-CoV-2 reference sequences of the nine recognized GISAID clades publically available in the GISAID database using MUSCLE multiple sequence alignment algorithms (32) implemented in MEGAX (33). Phylogenetic analyses were performed on nucleotide sequences using the maximum likelihood method with the Tamura 3-parameter model then on amino acid sequences, obtained from the aligned sequences, using the maximum likelihood method and the Jones Taylor Thornton model. The tree topologies were supported by 1000 bootstrap replicates.

Mutation profiles in the ORF1a, ORF1b, S, ORF3a, E, M, ORF6, ORF7a, ORF8, N, and ORF10 genomic regions of SARS-CoV-2 were assessed, by comparing the nucleotide and deduced amino acid sequences of the Tunisian strains with those of the Wuhan reference strain, using the sequence alignment performed by MUSCLE multiple sequence alignment algorithms (32) implemented in MEGAX (33).

Results

Phylogenetic tree in **Figure2** was performed based on the alignment of the 618-nucleotides fragment in the S gene of the 201 studied Tunisian SARS-CoV-2 strains, together with the 9 selected references SARS-CoV-2 sequences according to the GISAD nomenclature. The tree topology shows that the Tunisian sequences are divided into 3 different clusters. Cluster1, represented in purple color, includes the highest number of sequences (174 out of 201, 86.5% of Tunisian strains) that clustered with the 4 reference sequences of the GISAID Clades G, GH, GR and GV. The phylogenetic distribution within this cluster shows several phylogenetic sub-branches reflecting a large genetic variability. Cluster2 indicated in blue color comprises 15 identical sequences that clustered with the GISAID reference sequence from Clade GRY. Cluster3, indicated in red color, contains 12 sequences that clustered with the GISAID reference sequence from Clade S.

Eighteen representative samples from these clusters, indicated by a green square in Figure2, were selected for whole genome sequencing: 13 from Cluster1, 2 from Cluster2 and 3 from Cluster3. The phylogenetic tree of the obtained 18 whole genome sequences, together with the 9 GISAID references SARS-CoV-2 sequences, is shown in **Figure3**. The figure also shows the classification of the Tunisian sequences according to the PANGO and the Nextstrain classifications. The phylogenetic distribution of the sequences based on whole genome sequences (Figure3) is similar to the one obtained in Figure2, based on the partial S gene genomic data.

The 13 sequences from Cluster1 highlighted in purple color in Figure2 grouped together within PANGO B lineage in Figure3. The phylogenetic distribution of these sequences clearly shows the presence of 3 sub-clusters called sub-cluster 1a, 1b and 1c classified as clade G/20A, GV/20A-C and GH/20C respectively according to the GISAID/Nextstrain nomenclatures. Sub-cluster 1a is represented by only one sequence (SP-0362), while Sub-cluster 1b and Sub-cluster 1c are represented by 4 (SP-0202, SP-0083, SP-0377 and SP-0036) and 8 (SP-0378, SP-0017, SP-0382, SP-0210, SP-0084, SP-0089, SP-0055 and SP-0105) sequences respectively.

Two sequences from Cluster2 in Figure 2 were also found to cluster together with the reference sequence of the GR GISAID Clade based on whole genome sequencing comparison; the sequences also belong to the PANGO B lineage and to the 20B Clade of the Nextstrain nomenclature.

Unlike the sequences from Cluster1 and Cluster2, the three whole genome sequences from Cluster3 belong to the PANGO A lineage. They grouped together with the reference sequence of the S GISAID Clade, similarly to the results obtained based on the partial S sequences.

The amino acid sequences related to the 201 partial S sequences and the 18 whole genome sequences were deduced from the obtained nucleotide sequences and compared to the Wuhan reference protein sequences.

Table 1 shows the amino acid substitution profile in the sequenced fragment of the S gene of the 201 samples investigated in the present study. Fourteen different mutation profiles were found. Most of the sequences (147/174) had zero non synonymous mutation as compared to the Wuhan reference, excepting the D614G which was found in all the sequences from cluster 1 and Cluster 2. The remaining 27 sequences from Cluster1 had 1 to 2 additional substitutions within the sequenced fragment (**Table 1**). The 15 sequences from Cluster 2 shared an identical mutational profile with the amino acid substitutions N501Y, A570D, D614G and P681H which are known to be characteristic of the VOC B.1.1.7 initially detected in the UK. The 12 sequences from Cluster did not have the D614G substitution but three mutations that suggest the VUI A.27 (N501Y, A653V, Q655H); one sequence (SP-0347 which was in a separate branch within the phylogenetic tree shown in Figure2), had an additional substitution (Q677H).

Table2 shows the amino acid substitution profile along the whole genome of the 18 selected Tunisian SARS-CoV-2 and representative from the different clusters found based on S partial sequences. The two sequences from Cluster 2 had identical mutational profile in the S gene and a total of 23 and 24 amino acid substitution along the whole genome; these results confirm the belonging of the two sequences to the B1.1.7 lineage (VOC). The three sequences from Cluster 3 shared 15 identical amino acid substitutions along the whole genome and the results confirm the belonging of the three sequences to the A.27 lineage, identified as variant of interest (VOI) initially detected in France. Among Cluster 1, one sequence (SP062 – Sub-Cluster 1a) had a mutational profile that corresponds to the identified variant of interest (VOI) B.1.525 initially detected in Nigeria and in the UK. The sequences from Sub-Cluster 1c shared several identical mutations in the non structural regions of the genome and belonged to the B.1.160 lineage that is not presently identified as VOC or VOI. The same is for the sequences from Sub-Cluster 1b that had more genetic diversity and belonged to the B.1.177 lineage.

Discussion

Since the beginning of the COVID-19 pandemic, several SARS-CoV-2 variants emerged, some of them totally changed the infection epidemiology. First, a variant with the D614G mutation emerged and became dominant globally (34). In our series, this mutation is found in 186 out of the 201 isolates (92%). Other variants emerged subsequently and it is now hypercritical to track the already known of them labelled as VOCs or VOIs and also to monitor the emergence of new variants. The method of choice is whole genome sequencing using NGS high throughput technologies which improved considerably during the last years with a cost that declines continuously. However, despite this progress, NGS is still expensive and resources for this technology remain limited in many low and middle income countries where it is not possible to perform whole genome sequencing for representative number of SARS-CoV-2 positive cases. Thus, the use of other technologies to identify isolates that are to be sequenced in priority is highly needed. Several real-time PCR tests that target the already known VOCs, especially the B.1.1.7 (United Kingdom), B.1.351 (South Africa) and B1.1.28 (Brazil) are now commercially available. They can be very useful to rapidly identify the introduction of these VOCs to a country/region and to monitor their transmission. However, these kits cannot detect other variants of interest that already emerged, or that may emerge any time. Furthermore, other variants can be characterized by the failure to detect the S gene in these tests, known as S gene target failure (SGTF) (35).

In the present work, we developed a couple of primers allowing to generate a 618-nucleotide-long sequence in the viral S gene that includes key mutations of the already known VOCs and VOIs. Sequencing of this fragment by the traditional Sanger technology allows rapid identification of these VOCs and VOIs and helps to better select specimens that need to be sequenced by NGS technologies. Using this approach, it is possible to detect 14 amino acid substitutions that have been identified in several VOCs and VOIs (G482V, E484K, N501Y, A570D, D574Y, D614G, E619Q, A626S, D627E, A653V, H655Y, Q675H, Q677H and P681H) and to get a rapid orientation towards an already known or a new variant. In our series and using these primers, we were able to detect the first introduction of the B.1.1.7 (VOC) and two other VOIs (A.27 and B.1.525) and to select other viruses for WGS based on the results obtained in the S partial genomic region. The second step consisting in whole genome sequencing allowed to have a holistic view of all variants within the selected viral strains and confirmed the initial classification of the strains based on partial S gene sequencing.

The specimens included in the present work were collected in the decreasing phase of a COVID-19 wave that occurred in Tunisia starting from September 2020 up to January 2021. This period was characterized by a high transmission within the population and this explains the high genetic diversity that we found in the obtained sequences. Several lineages were identified and more than 100 different amino acid changes, in comparison to the standard Wuhan strain, were identified all through the viral genome.

During the study period, the first isolates of the VOC B.1.1.7, initially identified in the UK, were detected. The sequenced isolates had the H69del, V70del, Y144del, N501Y, A570D, D614G P681H, T716I, S982A, D1118H common amino acid substitutions with the 20I/501Y.V1 (UK variant). Thus, it is highly expected that the genetic features described herein will rapidly change to a lower genetic variability and a predominance of the B.1.1.7 UK lineage. Indeed, this is what happened in most countries of the world where the B.1.1.7 UK lineage was introduced causing devastating waves of COVID-19 (36, 37). With its higher transmissibility within the human population, it becomes rapidly predominant once introduced and this is also what is expected to happen in Tunisia.

Furthermore, we were able to detect viruses belonging to the A.27 lineage, initially detected in Danemark and now classified as VOI. This lineage was detected in around 26 different countries in the world, from Europe, Africa as well as USA and Australia.

Whole genome sequencing of three isolates in this series revealed the presence of amino acid substitutions characteristic of this lineage, including L18F, L452R, N501Y, A653V, H655Y, D796Y and G1219V and the absence of the D614G substitution in the Spike protein. One strain (SP-0347) presented two additional substitutions: P26S that is found in the P1 20J/501Y.V3 (Brazilian variant) and Q677H found in the *Henri Mondor variant* detected in different regions of France (38).

We have also detected one sequence (SP062 – Sub-Cluster 1a) with a mutational profile corresponding to the B.1.525, initially detected in Nigeria and in the UK. This variant was detected in 48 different countries in the world at writing time and is presently classified as VOI.

The rest of sequences from Sub-Clusters 1b and 1c belonged to the B.1.160 and B.1.177 lineages that are not presently identified as VOCs or VOIs. These sequences exhibit quite high genetic variability which is expected after the high active transmission period that the country experienced in late 2020 and January 2021. Among all these variants, some may then disappear and other may persist or even dominate if they have a selective advantage in terms of virulence or transmissibility.

Conclusion

In conclusion, this study gives an overview of the SARS-CoV-2 strains circulating in Tunisia after a high transmission wave of COVID-19. Partial S gene sequencing followed by whole genome sequencing of a selection of specimen was used to identify the different circulating variants. This strategy may be of interest for several countries; it helps to establish a genomic surveillance that is now highly needed in all regions of the world, with a good cost/effectiveness ratio.

ACKNOWLEDGMENTS

This work was co-founded by the Tunisian Ministry of High education and Research and the European Union’s Horizon 2020 research and innovation program under grant agreement No 883441, project STAMINA (Demonstration of intelligent decision support for pandemic crisis prediction and management within and across European borders). The authors thank the field staff from the Ministry of Health for their efforts in specimen collection and transportation to the laboratory.

References

1. Lu H, Stratton CW, Tang Y-W. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol. 2020/02/12 éd. avr 2020;92(4):401-2.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W; China Novel Coronavirus Investigating and Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020 Feb 20;382(8):727-733. doi: 10.1056/NEJMoa2001017. Epub 2020 Jan 24. PMID: 31978945; PMCID: PMC7092803.
3. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, Fabiani C, Anselmo A, Ciammaruconi A, Fortunato A, De Santis R, Fillo S, Capobianchi MR, Gismondo MR, Ciervo A, Rezza G, Castrucci MR, Lista F, On Behalf Of Iss Covid-Study Group. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. Euro Surveill. 2020 Apr;25(13):2000305. doi: 10.2807/1560-7917.ES.2020.25.13.2000305. PMID: 32265007; PMCID: PMC7140597.
4. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. N Engl J Med [Internet]. 30 mars 2020 [cité 23 mars 2021]; Disponible sur: https://www.nejm.org/doi/10.1056/NEJMp2005630
5. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020 Jan 28;9(1):221-236. doi: 10.1080/22221751.2020.1719902. Erratum in: Emerg Microbes Infect. 2020 Dec;9(1):540. PMID: 31987001; PMCID: PMC7067204.
6. Yuan M, Wu NC, Zhu X, Lee CD, So RTY, Lv H, Mok CKP, Wilson IA. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science. 2020 May 8;368(6491):630-633. doi: 10.1126/science.abb7269. Epub 2020 Apr 3. PMID: 32245784; PMCID: PMC7164391.

7. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260-1263. doi: 10.1126/science.abb2507. Epub 2020 Feb 19. PMID: 32075877; PMCID: PMC7164637.
8. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 16 avr 2020;181(2):281-292.e6.
9. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI; Sheffield COVID-19 Genomics Group, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020 Aug 20;182(4):812-827.e19. doi: 10.1016/j.cell.2020.06.043. Epub 2020 Jul 3. PMID: 32697968; PMCID: PMC7332439.
10. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, Southgate J, Johnson R, Jackson B, Nascimento FF, Rey SM, Nicholls SM, Colquhoun RM, da Silva Filipe A, Shepherd J, Pascall DJ, Shah R, Jesudason N, Li K, Jarrett R, Pacchiarini N, Bull M, Geidelberg L, Siveroni I; COG-UK Consortium, Goodfellow I, Loman NJ, Pybus OG, Robertson DL, Thomson EC, Rambaut A, Connor TR. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 2021 Jan 7;184(1):64-75.e11. doi: 10.1016/j.cell.2020.11.020. Epub 2020 Nov 19. PMID: 33275900; PMCID: PMC7674007.
11. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, Farzan M, Choe H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv [Preprint]. 2020 Jun 12:2020.06.12.148726. doi: 10.1101/2020.06.12.148726. Update in: Nat Commun. 2020 Nov 26;11(1):6013. PMID: 32587973; PMCID: PMC7310631.
12. Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, Slayton RB, Tong S, Silk BJ, Armstrong GL, Biggerstaff M, Dugan VG. Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29,

2020-January 12, 2021. MMWR Morb Mortal Wkly Rep. 2021 Jan 22;70(3):95-99. doi: 10.15585/mmwr.mm7003e2. PMID: 33476315; PMCID: PMC7821772.

13. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill. 2021 Jan;26(1):2002106. doi: 10.2807/1560-7917.ES.2020.26.1.2002106. Erratum in: Euro Surveill. 2021 Jan;26(3): PMID: 33413740; PMCID: PMC7791602.
14. Makoni M. South Africa responds to new SARS-CoV-2 variant. Lancet. 2021 Jan 23;397(10271):267. doi: 10.1016/S0140-6736(21)00144-6. PMID: 33485437; PMCID: PMC7825846.
15. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 2021 Apr;592(7854):438-443. doi: 10.1038/s41586-021-03402-9. Epub 2021 Mar 9. PMID: 33690265.
16. Voloch CM, da Silva Francisco R Jr, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, Guimarães APC, Mariani D, da Costa RM, Ferreira OC Jr; Covid19-UFRJ Workgroup, LNCC Workgroup, Adriana Cony Cavalcanti, Frauches TS, de Mello CMB, Leitão IC, Galliez RM, Faffe DS, Castiñeiras TMPP, Tanuri A, de Vasconcelos ATR. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. J Virol. 2021 Mar 1:JVI.00119-21. doi: 10.1128/JVI.00119-21. Epub ahead of print. PMID: 33649194.
17. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv. 1 janv 2021;2020.12.30.20249034.
18. Kemp S, Harvey W, Lytras S, Carabelli A, Robertson D, Gupta R. Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion H69/V70. bioRxiv. 1 janv 2021;2020.12.14.422555.

19. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. bioRxiv. 19 nov 2020;2020.11.19.389916.

20. Washington NL, White S, Barrett KMS, Cirulli ET, Bolze A, Lu JT. S gene dropout patterns in SARS-CoV-2 tests suggest spread of the H69del/V70del mutation in the US. medRxiv. 30 déc 2020;2020.12.24.20248814.

21. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Luchsinger L, Hillyer CD, Caskey M, Robbiani DF, Rice CM, Nussenzweig MC, Hatziioannou T, Bieniasz PD. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife. 2020 Oct 28;9:e61312. doi: 10.7554/eLife.61312. PMID: 33112236; PMCID: PMC7723407.

22. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe. 2021 Mar 10;29(3):463-476.e6. doi: 10.1016/j.chom.2021.02.003. Epub 2021 Feb 8. PMID: 33592168; PMCID: PMC7869748.

23. Saidi O, Malouche D, Saksena P, Arfaoui L, Talmoudi K, Hchaichi A, Bouguerra H, Romdhane HB, Hsairi M, Ouhichi R, Souteyrand Y, Alaya NB; NONED Working Group. Impact of contact tracing, respect of isolation and lockdown in reducing the number of cases infected with COVID-19: Case study: Tunisia's response from March 22 to 04 May 2020. Int J Infect Dis. 2021 Feb 9:S1201-9712(21)00096-5. doi: 10.1016/j.ijid.2021.02.010. Epub ahead of print. PMID: 33578008; PMCID: PMC7872851.

24. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill ;25(3). doi: 10.2807/1560-7917.ES.2020.25.3.2000045.

25. Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, Ng DYM, Wan CKC, Yang P, Wang Q, Peiris M, Poon LLM. Molecular Diagnosis of a Novel

Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clin Chem. 2020 Apr 1;66(4):549-555. doi: 10.1093/clinchem/hvaa029. PMID: 32031583; PMCID: PMC7108203.

26. Yoon H, Leitner T. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics. 1 mai 2015;31(9):1472.4.

27. Brodin J, Krishnamoorthy M, Athreya G, Fischer W, Hraber P, Gleasner C, Green L, Korber B, Leitner T. A multiple-alignment based primer design algorithm for genetically highly variable DNA targets. BMC Bioinformatics. 2013 Aug 21;14:255. doi: 10.1186/1471-2105-14-255. PMID: 23965160; PMCID: PMC3765731.

28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.

29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455-77. doi: 10.1089/cmb.2012.0021. Epub 2012 Apr 16. PMID: 22506599; PMCID: PMC3342519.

30. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017 Jan 10;1(1):33-46. doi: 10.1002/gch2.1018. PMID: 31565258; PMCID: PMC6607375.

31. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017 Mar 30;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494. PMID: 28382917; PMCID: PMC5388101.

32. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792-7. doi: 10.1093/nar/gkh340. PMID: 15034147; PMCID: PMC390337.

33. Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

34. Isabel S, Graña-Miraglia L, Gutierrez JM, Bundalovic-Torma C, Groves HE, Isabel MR, Eshaghi A, Patel SN, Gubbay JB, Poutanen T, Guttman DS, Poutanen SM. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. Sci Rep. 2020 Aug 20;10(1):14031. doi: 10.1038/s41598-020-70827-z. PMID: 32820179; PMCID: PMC7441380.

35. Bal A, Destras G, Gaymard A, Stefic K, Marlet J, Eymieux S, Regue H, Semanas Q, d'Aubarede C, Billaud G, Laurent F, Gonzalez C, Mekki Y, Valette M, Bouscambert M, Gaudy-Graffin C, Lina B, Morfin F, Josset L; COVID-Diagnosis HCL Study Group. Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69-V70, France, August to December 2020. Euro Surveill. 2021 Jan;26(3):2100008. doi: 10.2807/1560-7917.ES.2021.26.3.2100008. PMID: 33478625; PMCID: PMC7848679.

36. Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, Petrone ME, Breban MI, Watkins AE, Vogels CBF, Kalinich CC, Dellicour S, Russell A, Kelly JP, Shudt M, Plitnick J, Schneider E, Fitzsimmons WJ, Khullar G, Metti J, Dudley JT, Nash M, Beaubier N, Wang J, Liu C, Hui P, Muyombwe A, Downing R, Razeq J, Bart SM, Grills A, Morrison SM, Murphy S, Neal C, Laszlo E, Rennert H, Cushing M, Westblade L, Velu P, Craney A, Cong L, Peaper DR, Landry ML, Cook PW, Fauver JR, Mason CE, Lauring AS, St George K, MacCannell DR, Grubaugh ND. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell. 2021 May 13;184(10):2595-2604.e13. doi: 10.1016/j.cell.2021.03.061. Epub 2021 Apr 3. PMID: 33891875; PMCID: PMC8018830.

37. Loconsole D, Sallustio A, Accogli M, Leaci A, Sanguedolce A, Parisi A, Chironna M. Investigation of an outbreak of symptomatic SARS-CoV-2 VOC 202012/01-lineage B.1.1.7 infection in healthcare workers, Italy. Clin Microbiol Infect. 2021 May 10:S1198-743X(21)00228-7. doi: 10.1016/j.cmi.2021.05.007. Epub ahead of print. PMID: 33984489; PMCID: PMC8107058.

38. Fourati S, Decousser JW, 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, Pawlotsky JM, Rodriguez C. Novel SARS-CoV-2 Variant Derived from Clade 19B, France. Emerg Infect Dis. 2021

May;27(5):1540-1543. doi: 10.3201/eid2705.210324. PMID: 33900195; PMCID: PMC8084519.

LEGENDS

Figure1. Samples collection period investigated in the present study

The graph displays the number of cases and the number of deaths in Tunisia since the declaration of the pandemy in March 2020. The Abscisse axe represents the number of weeks from March 2020 till May 2021. Weeks highlighted in red color represents the samples collection period investigated in the present study.

Figure2. Phylogenetic tree of 201 SARS-CoV-2 sequences based on partial S gene nucleotide sequencing.

The phylogenetic tree includes 201 Tunisian sequences compared to 9 representative reference sequences of SARS-Cov-2 Clades. The tree was performed using the neighbor joining method and the Tamura 3-parameter (T92) model. Topology was supported by 1000 bootstrap replicates. The sequences reported in this study are shown in bold, and indicated by the laboratory code. The sequences downloaded from GISAID are indicated by their accession number followed. Cluster 1 in purple color denotes sequences presenting the D614G substitution and the lack of the amino acid substitution N501Y. Cluster 2 in blue color includes sequences having the N501Y, A570D, D614G and P681H substitutions. Cluster 3 in red color groups sequences with the N501Y, A653V and H655Y substitutions and the lack of the amino acid substitution D614G.

Figure3. Phylogenetic tree of 18 SARS-CoV-2 whole genome sequences circulating in Tunisia compared to 9 reference strain genomes.

The phylogenetic tree includes 18 Tunisian sequences compared to 9 representative reference sequences of SARS-Cov-2 Clades. The tree was performed using the neighbor joining method and the Tamura 3-parameter (T92) model. Topology was supported by 1000 bootstrap replicates. The sequences reported in this study are shown in bold, and indicated by the laboratory code. The sequences downloaded from GISAID database are indicated by their accession number. Cluster 1 in purple color, Clade 2 in blue color and Clade 3 in red.

Table.1 Amino acid substitution profile in the sequenced fragment of the S gene

Mutation profile	Cluster.1 N=174	Cluster.2 N=15	Cluster.3 N=12
E484K, D614G	02	0	0
E484K, D614G, Q677H	01	0	0
D614G, S637L, A647S	01	0	0
D614G, I666L	01	0	0
D614G, Q675L	01	0	0
D574Y, D614G, A626S	01	0	0
D614G, A626S	02	0	0
D614G, V622F	01	0	0
D614G, E619Q	01	0	0
D614G, D627E	16	0	0
D614G	147	0	0
N501Y, A570D, D614G, P681H	0	15	0
N501Y, A653V, Q655H	0	0	11
N501Y, A653V, Q655H, Q677H	0	0	01

Table 2: Mutation profile of the 18 obtained Sars-Cov-2 Tunisian strains

Cluster 2			
Gene	mutation	SP-0393	SP-0343
NSP3	T183I	▨	▨
	A890D	▨	▨
	I1412T		▨
NSP5	K90R		▨
NSP6	del S106		▨
	del G107		▨
	del F108		▨
NSP12	P227L	▨	
	P323L	▨	▨
NSP13	A237V	▨	
S	del H69	▨	▨
	del V70	▨	▨
	del Y144	▨	▨
	N501Y	▨	▨
	A570D	▨	▨
	D614G	▨	▨
	P681H	▨	▨
	T716I	▨	▨
	S982A	▨	▨
	D1118H	▨	▨
NS3ab	Q57H	▨	▨
NS8	Q27 STOP	▨	▨
	K68 STOP	▨	
	Y73C	▨	▨
N	D3L	▨	▨
	R203K	▨	▨
	G204R	▨	▨
	S235F	▨	▨

VOC

20I/501Y,V1 (B.1.1.7)

Clade GRY

Cluster 3				
Gene	mutation	SP-0154	SP-0157	SP-0347
NSP2	P106L	▨	▨	▨
	P106L	▨	▨	▨
NSP3	L368?			▨
NSP4	D217G	▨	▨	▨
	T319I	▨		
NSP5	S123F	▨		
NSP6	L37F			▨
	N82S	▨	▨	▨
NSP9	P57S	▨		
NSP13	P77L	▨	▨	▨
	P491S			▨
S	L18F	▨	▨	▨
	P26S			▨
	V227A	▨		
	L452R	▨	▨	▨
	N501Y	▨	▨	▨
	A653V	▨	▨	▨
	H655Y	▨	▨	▨
	Q677H			▨
	D796Y	▨	▨	▨
	K1191N	▨	▨	
	G1219V	▨	▨	▨
NS3ab	V50A	▨	▨	▨
	G172C		▨	
NS8	A65S		▨	▨
	L84S	▨	▨	▨
N	S202N	▨	▨	▨

VOI

19B/501Y (A.27)

Clade S

Sub-Cluster 1a		
Gene	mutation	SP-0362
NSP3	T1189I	▨
	K1693N	▨
NSP6	del S106	▨
	del G107	▨
	del F108	▨
NSP9	T21I	▨
	T109I	▨
NSP12	P323L	▨
S	Q52R	▨
	A67V	▨
	del H69	▨
	del V70	▨
	del Y144	▨
	E484K	▨
	D614G	▨
	Q677H	▨
	F888L	▨
E	L21F	▨
M	I82T	▨
NS8	A12G	▨
	T205I	▨

VOI

20A/484K (B.1.525)

Clade G

Sub-Cluster 1c									
Gene	mutation	SP-0202	SP-0378	SP-0382	SP-0017	SP-0084	SP-0105	SP-0089	SP-0055
NSP2	E57A				▨				
	T153A					▨			
	T497I							▨	
	M551I							▨	
	K618N	▨							
NSP4	M324I	▨	▨	▨	▨	▨	▨	▨	▨
	L438I	▨	▨	▨	▨	▨	▨	▨	▨
NSP12	A185S	▨	▨	▨	▨	▨	▨	▨	▨
	P323L	▨	▨	▨	▨	▨	▨	▨	▨
	V776L	▨	▨	▨	▨	▨	▨	▨	▨
NSP13	K218R	▨	▨	▨	▨	▨	▨	▨	▨
	E261D	▨	▨	▨	▨	▨	▨	▨	▨
	M429I					▨	▨	▨	▨
NSP14	P43S							▨	
	S450R								▨
	T516I		▨						
NSP15	A94V	▨							
	V127S							▨	
	E202G	▨		▨	▨				
	P262S					▨			
S	S477N	▨	▨	▨	▨	▨	▨	▨	▨
	A574Y							▨	
	D614G	▨	▨	▨	▨	▨	▨	▨	▨
	A626S						▨	▨	▨
	D627E	▨		▨	▨				
	P812R		▨						
	A1020S	▨		▨	▨				
	K1157N						▨		
NS3ab	Q57H	▨	▨	▨	▨	▨	▨	▨	▨
NS7a	P99L	▨							
NS8	Q18 STOP						▨		
	S54L		▨						
	V62L	▨		▨					
	L95F		▨						
	S186P		▨						
	M234I	▨	▨		▨	▨	▨	▨	▨
	A376T	▨	▨		▨	▨	▨	▨	▨

Lineage B.1.160

Clade GH

Sub-Cluster 1b					
Gene	mutation	SP-0377	SP-0036	SP-0083	SP-0210
NSP2	E57A		▨		
NSP3	K429N	▨		▨	▨
NSP4	M324I		▨		
	L438I		▨		
NSP6	A54S	▨		▨	▨
NSP8	T141M			▨	
NSP12	A185S		▨		
	P323L	▨	▨	▨	▨
	A449V				▨
	V776L		▨		
NSP13	K218R		▨		
	E261D		▨		
	T366M				▨
	A505T	▨			
NSP15	E202G		▨		
S	P9S			▨	
	D138Y	▨			
	A222V	▨		▨	▨
	S477N		▨		
	D614G	▨	▨	▨	▨
	E619Q				▨
	V622F	▨			
	D627E		▨		
	Q675H			▨	
	T859I			▨	
	A1020S		▨		
	K1038E				▨
NS3ab	Q57H		▨		
M	A38S			▨	
NS7b	D36A			▨	
	A220V	▨		▨	▨
	Q229H			▨	
	M234I		▨		
	A376T		▨		

Lineage B1.177

Clade GV

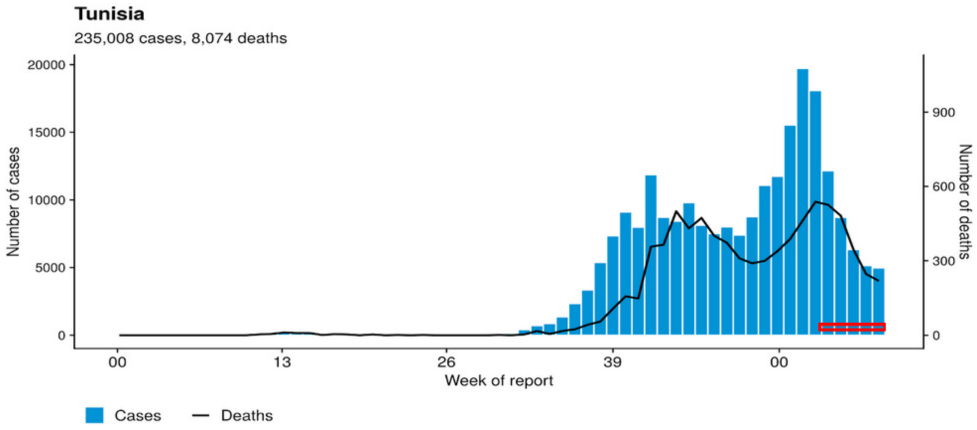


Figure1: Samples collection period investigated in the present study

Figure.2

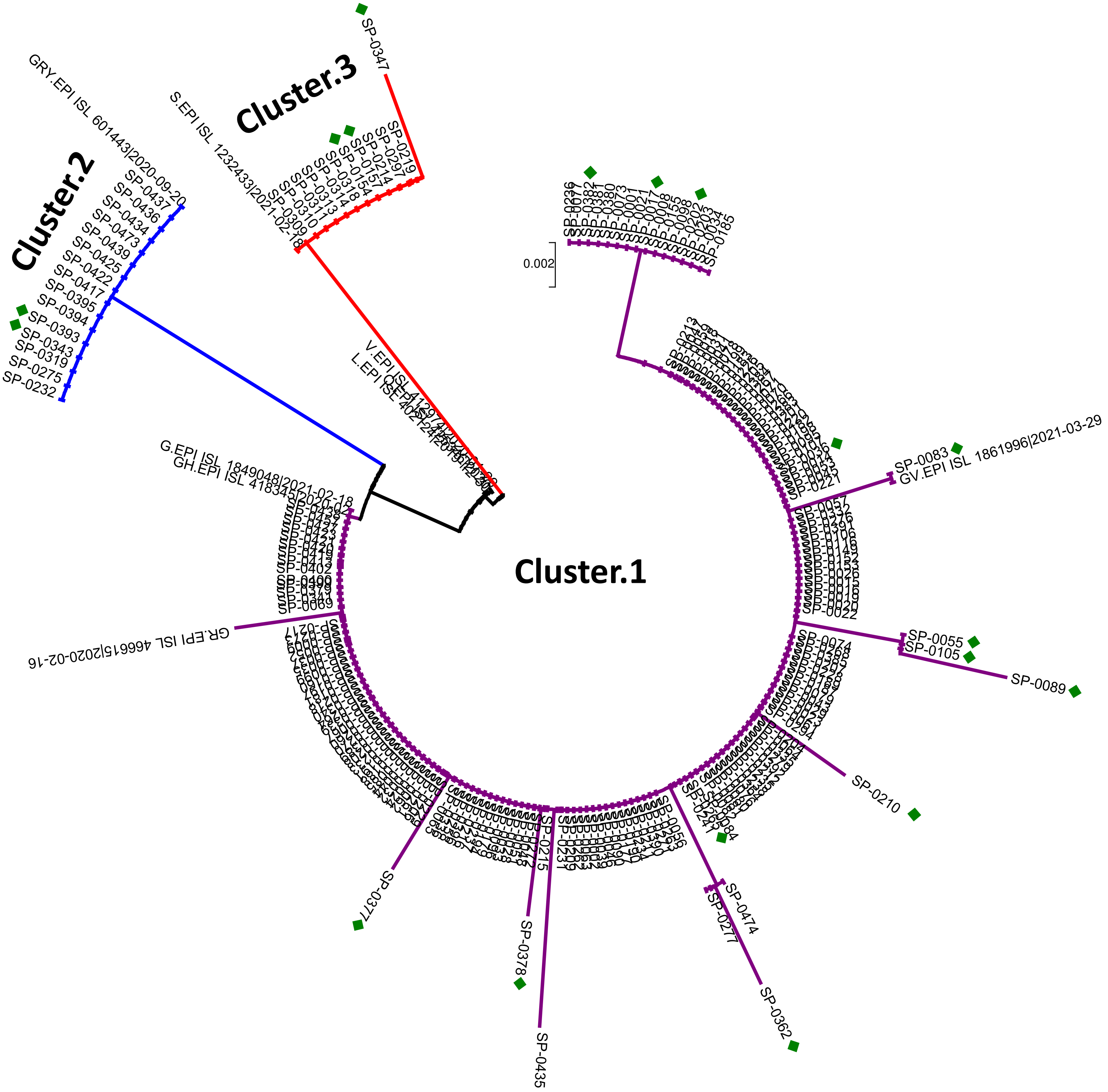
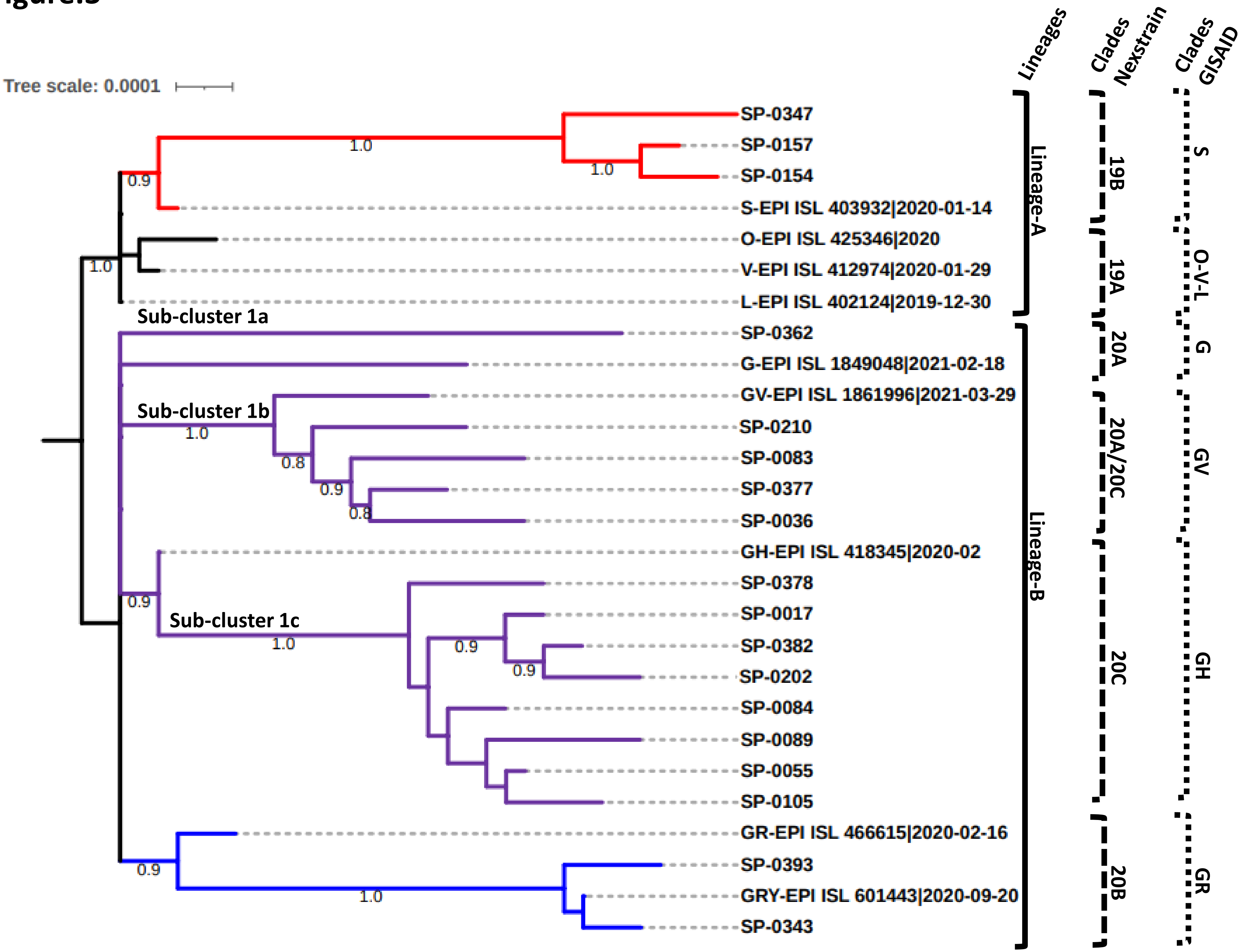

Cluster.1
Cluster.2
Cluster.3
0.002
GRY.EPI ISL 601443|2020-09-20
S.EPI ISL 1232433|2021-02-18
V.EPI ISL
L.EPI ISL
G.EPI ISL 1849048|2021-02-18
GH.EPI ISL 418345|2020-0
GR.EPI ISL 466615|2020-02-16
GV.EPI ISL 1861996|2021-03-29
SP-0347
SP-0219
SP-0297
SP-0214
SP-0157
SP-0154
SP-0318
SP-0313
SP-0437
SP-0436
SP-0434
SP-0473
SP-0439
SP-0425
SP-0422
SP-0417
SP-0395
SP-0394
SP-0393
SP-0343
SP-0319
SP-0275
SP-0232
SP-0236
SP-0185
SP-0221
SP-0083
SP-0057
SP-0022
SP-0055
SP-0105
SP-0089
SP-0074
SP-0210
SP-0474
SP-0277
SP-0362
SP-0215
SP-0378
SP-0435
SP-0377
SP-0402
SP-0069

Figure.3



Tree scale: 0.0001
Lineages
Clades Nexstrain
Clades GISAID
SP-0347
SP-0157
SP-0154
S-EPI ISL 403932|2020-01-14
O-EPI ISL 425346|2020
V-EPI ISL 412974|2020-01-29
L-EPI ISL 402124|2019-12-30
SP-0362
G-EPI ISL 1849048|2021-02-18
GV-EPI ISL 1861996|2021-03-29
SP-0210
SP-0083
SP-0377
SP-0036
GH-EPI ISL 418345|2020-02
SP-0378
SP-0017
SP-0382
SP-0202
SP-0084
SP-0089
SP-0055
SP-0105
GR-EPI ISL 466615|2020-02-16
SP-0393
GRY-EPI ISL 601443|2020-09-20
SP-0343
Sub-cluster 1a
Sub-cluster 1b
Sub-cluster 1c
1.0
1.0
0.9
1.0
1.0
0.8
0.9
0.8
0.9
1.0
0.9
0.9
0.9
1.0
Lineage-A
Lineage-B
19B
19A
20A
20A/20C
20C
20B
S
O-V-L
G
GV
GH
GR