Available online
https://www.atrss.dz/ajhs

Original Article

Analyse protéomique du SARS-CoV-2 responsable du COVID-19 en Algérie (étude in silico)

Proteomic analysis of SARS-CoV-2 responsible for COVID-19 in Algeria (in silico study)

***Boudjema Abdallah*[1], *Haouhach Sadika*[2], *Azzoune Asmaa*[1], *Louhibi Lotfi*[1].**

[1]*Molecular and Cellular Genetics Laboratory, University of Sciences and Technology of Oran "Mohamed Boudiaf", Oran, Algeria.*

[2] *Department of Biotechnology, Faculty of Natural and Life Sciences, University of Sciences and Technology of Oran "Mohamed Boudiaf", Oran, Algeria.*

ABSTRACT

The high variability of the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) genome presents a challenge to scientists on molecular diagnosis, treatment and vaccination. In this work, we have characterized the proteomic mutations of SARS-CoV-2 responsible for COVID-19 in Algeria present during the period March 2020 to July 2021. We extracted 44 sequences from the GISAID platform's database (Global Initiative on Sharing Avian Influenza Data). The search for virus variants was performed on the GISAID and PANGO platforms (Global Outbreak: cov-lineages.org/). The proteins with the most mutations were Spike, N (Nucleoprotein), NSP3 (Non-Structural Protein 3) and NSP12 (Non-Structural Protein 12). The most frequent mutations were D614G and P323L, found in the Spike and NSP12 proteins, respectively. However, positions S-614 and NSP12-323 were less mutable than those of the accessory NS7a/b proteins, which are involved, in the immune escape. Among the mutations found in the Spike, five were associated with the VOCs "Variants Of Concern" and four with the VOIs "Variants Of Interest".
The study of the GISAID clades revealed that the G, GH and GR clades were present at the start of the pandemic and replaced by the GK corresponding to the Delta variant, which is currently predominant in Algeria.

KEYWORDS : SARS-CoV-2, proteome, mutations, variants, Algeria.

RÉSUMÉ

La variabilité importante du génome de SARS-CoV-2 (Severe Acute Respiratory Syndrome CoronaVirus 2) pose de nombreux défis aux chercheurs tant au niveau du diagnostic moléculaire, du traitement que dans la vaccination des populations. Dans ce travail nous avons caractérisé les mutations du SARS-CoV-2 responsable du COVID-19 en Algérie présents durant la période mars 2020 à juillet 2021. Nous avons extrait 44 séquences complètes de la base de données de la plateforme GISAID (Global Initiative on Sharing Avian Influenza Data : www.gisaid.org). La recherche des variants du virus a été réalisée sur la plateforme PANGO (Global Outbreak : cov-lineages.org/) .Les protéines les plus mutées sont S (Spike), N (Nucléoprotéine), NSP3 (Protéine Non-Structurale 3) et NSP12 (Protéine Non-Structurale 12). Les mutations les plus fréquentes étaient D614G et P323L trouvées dans les protéines S et NSP12 respectivement. Cependant les positions S-614 et NSP12-323 étaient moins mutables que celles des protéines accessoires NS7a/b, qui

sont impliquées, dans l'échappement immunitaire. Parmi les mutations trouvées au niveau de la protéine S, cinq étaient associées aux variants VOCs « Variants Of Concern » et quatre avec les variants VOIs « Variants Of Interest ».
Par ailleurs, l'étude des clades GISAID a révélé que les clades G, GH, GR étaient présents au début de la pandémie puis progressivement ils ont été remplacés par le clade GK correspondant au variant delta (B.1.617.2) qui est actuellement prédominant en Algérie.

MOTS CLES: SARS-CoV-2, protéome, mutations, variants, Algérie.

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mailAddress: abdallah.boudjema@gmail.com

Received on: 28/02/2022
Revised on: 21/03/2022
Accepted on: 26/04/2022

DOI : 10.5281/zenodo.6491319

Introduction

In December 2019, Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) was identified as the viral agent responsible for the COVID-19 pandemic. In September 2021, the number of cases worldwide was over 93 million and 2 million deaths (World Health Organization, https://www.who.int).

SARS-CoV-2 belongs to the Coronaviridae family; it is a positive sense single-stranded RNA virus with a genome of approximately 30,000 nucleotides.
On the 5' side of the SARS-CoV-2 genome, two open reading frames ORF1a and 1b encode 16 non-structural proteins (NSP1 to NSP16) and at least 6 accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF10).Four structural proteins were encoded by this genome: glycoprotein or Spike (S), envelope protein (E), membrane glycoprotein (M) and nucleocapecid protein (N) [1-2].
SARS-CoV-2, like all RNA viruses, has a very high mutation rate, up to a million times that of its hosts [3]. This high rate is correlated with a modulation of virulence and with adaptability, which gives the virus the possibility of propagation. Mutation hot spots are gradually being identified in the SARS-CoV-2 genome. These sites are numerous and distributed in several regions of the genome.

A whole proteome analysis of SARS-CoV-2 revealed the presence of 7, 28, and 2081 mutations at the start of the pandemic (from December 2019 to May 2020), with a high frequency (10%), moderate (1– 10%), and weakly (0.01–1%) frequency, respectively. S: D614G (mutation in the spike) and NSP12: P323L (mutation in the NSP12 protein) were the most common mutations, occurring in nearly 3/4 of the cases. Furthermore, the envelope proteins (E) and the Nsp4, Nsp9and Nsp10 proteins were the least variable [4].

The accumulation of mutations in the Spike of the original virus from Wuhan (China), defined as variants called "Variants Of Interest" (VOIs) and "Variants Of Concern" (VOCs).

The VOCs variants are characterized by the presence of the D614G, N501Y, L452R substitutions and HV69-70del deletions in Spike. Indeed, the World Health Organization (WHO) has described four other VOCs: Alpha, Beta, Gamma and Delta.
The VOIs variants are characterized, in addition to the D614G, by the N439K, S477N, L452Q, and del247-252 and F490S mutations in the Spike. WHO also describes seven (07) VOIs variants: Epsilon, Zeta, Eota, P.3, Iota, Kappa and Lambda. These variants induce changes in the characteristics of the virus, in particular the affinity for the host cell receptor Angiotensin-converting enzyme 2 (ACE2) and antibody binding [5]. VOIs are associated with increased transmissibility and virulence; VOCs have, in addition to the attributes of VOIs, reduced efficiency of diagnostics, vaccines and treatments [6].

This exponential variability of the SARS-CoV-2 genome presents many challenges for scientists, including ensuring molecular diagnosis, treatment, and vaccination of populations, as well as adjusting their actions in response to the virus's evolution. The objective of this study was to characterize the proteomic mutations of SARS-CoV-2 responsible for COVID-19 in Algeria during the period March 2020 to July 2021 and to provide scientists and health authorities with molecular genetic data that would contribute to the effective management of the effects of the COVID-19 pandemic in Algeria.

Material and Methods

Material

The sequences of the SARS-CoV-2 virus explored in this study are those of the virus circulating in Algeria from March 2020 to July 2021. These sequences were extracted from the GISAID database (Global Initiative on Sharing Avian Influenza Data, www.gisaid.org). We studied 44 complete sequences of the SARS-CoV-2 virus genome (approximately 30,000 bp). In addition, for each sequence, further information was extracted

(date of collection, date of submission, age and sex of the patient).
The search for variations of the virus proteome was performed with reference to the sequence "hCoV-19/Wuhan/WIV04/2019".

Methods

The search for mutations in the virus proteome was made for each sequence on the GISAID platform (www.gisaid.org) with "Covsurver" (corona.bii.a-star.edu.sg). Apart from mutations, additional data is available (GISAID clades, geographical distribution and other statistics).
In this study, only mutations present at least three times are taken into account, they are considered non-random [4].
Percent mutations per protein (PMP) was calculated by counting mutations for each protein (PMP = number of amino acids mutated x 100 / total mutations), thus the minimum frequency was 3/ total mutations.
The recurrence of mutations is defined by their redundancy in the same position in the protein. In this investigation, the number of sequences studied was 44, so the minimum frequency of a recurrent mutation was 6.8% (3/44). Thus, we classified the mutations into three categories: recurrent (HR), recurrent (MR) and weakly recurrent (WR). A mutation was considered highly or moderately recurrent when occurring at a percentage ≥10% or ≥6.81 and<10%, respectively. It was considered weakly if its frequency is ≤6.81%.
Regarding the calculation of the frequency of a mutable position per protein (FMPP), it was performed by dividing the number of times the position was found to mutate by the amino acid length of the protein (FMPP = number of mutations at the same position/length of the protein) [4]. We only considered positions that have a 1% chance of being mutated into the protein of interest.
The search for virus variations was performed on the PANGO platform (Global Outbreak, cov-lineages.org/) with "Lineage Assigner". Variants are indicated by a letter (A, B, C...) and numbers punctuated by points.
Several mutations are found simultaneously in the genome of a viral strain. These mutations cause the variability of the genome and generate variants of the virus. These variations are positioned, according to their sequence similarities and relative to the Wuhan (China) reference sequence, in GISAID of the phylogenetic tree (www.gisaid.org). There are currently seven major clades: G, GH, GR, GK, S, V and L. An eighth O clade (designates strains that do not belong to any of the seven major clades). The temporal distribution of GISAID clades during the COVID-19 pandemic in Algeria was performed by counting the clades present each month from March 2020 to July 2021.

Results

Distribution of mutations and their recurrences in virus genome

In total, we found 818 mutations for the 44 sequences studied. The mutations were substitutions, nonsense mutations and deletions.
Of the 26 proteins of the virus 20 had mutations: these proteins were classified according to the frequency of mutations : S (27.87%)> N (15.04%)> NSP3 (10.15%)> NSP12 (9.66%)> NS7a (5.5%)> NSP6 (5.01%)> NSP4 (4.89%)> NS3 (4.66%)> NSP14 (3.18%)> NSP13 (2.81%)> M (2 , 45%)> NS8 (2.32%)> NS7b (2.2%)> NSP2 (2.08%)> NSP5 (0.61%), NSP16 (0.61%)> NSP1 (0.4 %), NSP15 (0.4%)> NSP10 (0.1%), NSP8 (0.1%) (Figures 1).
Six proteins (NSP7, NSP9, NSP11, ORF6, ORF10 and E) showed no mutations.

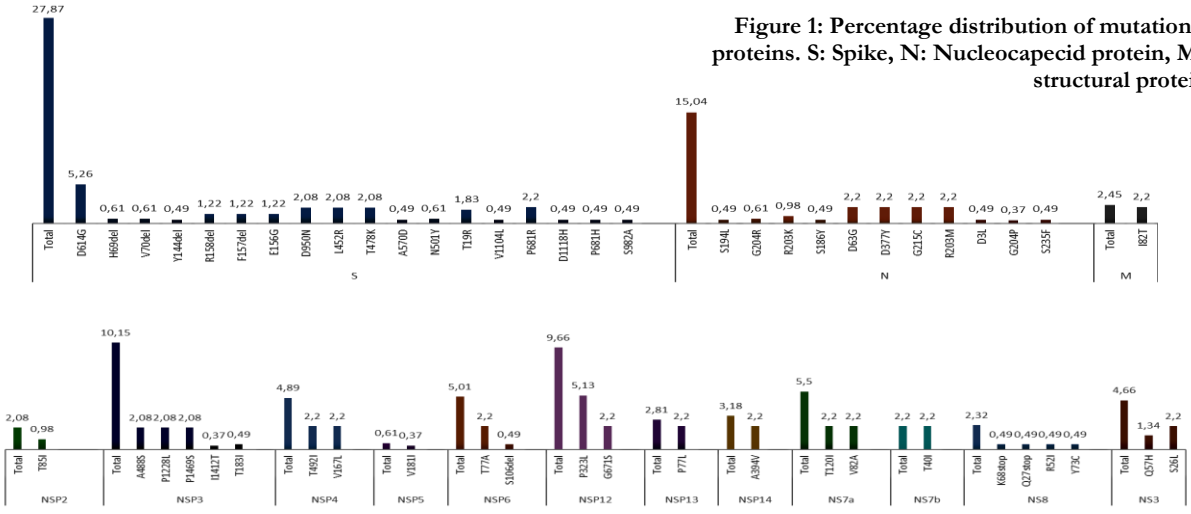

Figure 1: Percentage distribution of mutations (at least 3 variati proteins. S: Spike, N: Nucleocapecid protein, M: Membrane, NSP structural protein (accessory prote

Considering the minimum PMP of 0.36% (3/818), the mutations were classified according to the following distribution (Figure 1): S-D614G (5.26%)> NSP12-P323L (5.13%)> N- D63G (2.2%), N- D377Y (2.2%), N-G215C (2.2%), N-R203M (2.2%), S-P681R (2.2%), M-I82T (2.2%), NSP4-T492I (2.2%), NSP4-V167L (2.2%), NSP6-T77A (2.2%), NSP12-G671S (2.2%), NSP13- P77L (2.2%), NSP14-A394V (2.2%), NS7a-T120I (2.2%), NS7a-V82A (2.2%), NS7b-T40I (2.2%), NS3- S26L (2.2%)> S-D950N (2.08%), S-L452R (2.08%), S-T478K (2.08%), NSP3-A488S (2.08%), NSP3- E869D (2.08%), NSP3-P1228L (2.08%)>S-T19R (1.83%)> NS3-Q57H (1.34%)> S-R158del (1.22%), S- F157del (1.22%), S-E156G (1.22%)> N-R203K (0.98%), NSP2-T85I (0.98%)> S-H69del (0.61%), S- V70del (0.61%), S-N501Y (0.61%), N-G204R (0.61%)> S-Y144del (0.49%), S-A570D (0.49%), S-V1104L (0.49%), S-P681H (0.49%), S-S982A (0.49%), S-D1118H (0.49%), N-S194L (0.49%), N- S186Y (0.49%), N-D3L (0.49%), N-S235F (0.49%), NSP3-T183I (0.49%), NSP6-S106del (0.49%), NS8-K68stop (0.49%), NS8-Q27sto p (0.49%), NS8-R52I (0.49%), NS8-Y73C (0.49%)> N-G204P (0.37%), NSP3-I1412T (0.37%), NSP5- V181I (0.37%).

In addition, we discovered 35 highly recurrent mutations (>10%), two of which, S-D614G and NSP12-P323L, are found in 97% of the sequences examined. The rest of the highly recurrent mutations are localized in the proteins S, M, N, NSP12, NSP13 and NSP14 with an average frequency of 40%. Twenty-two mutations were moderately recurrent (≥6.81% and <10%) and four mutations were weakly recurrent (≤6.81%) (Figure 2).

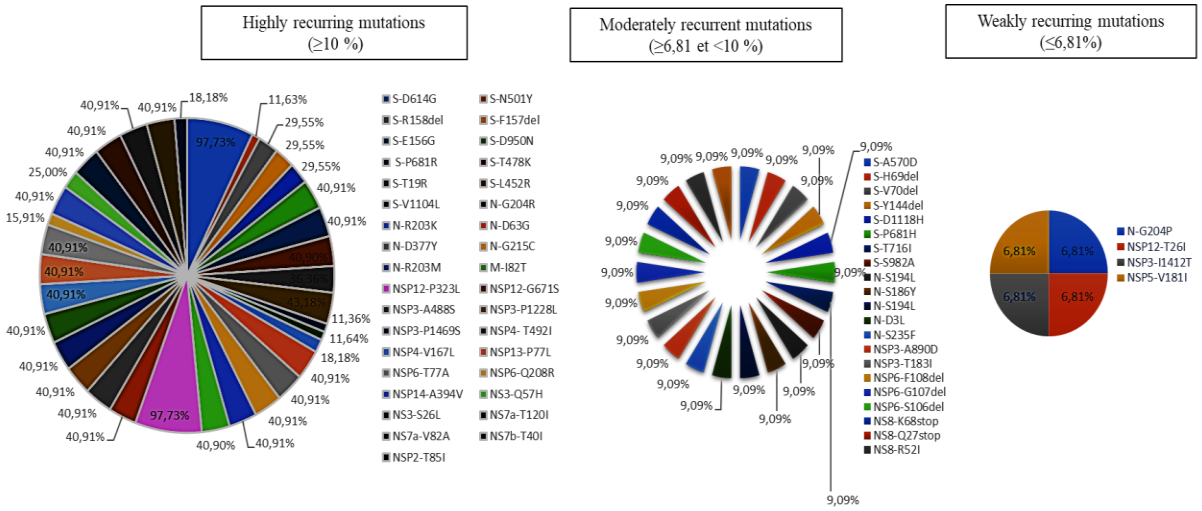

Figure 2: Percentage distribution of mutations according to their recurrences in SARS-CoV-2 sequences. S: Spike, N: Nucleocapecid protein, M: Membrane, NSP: Non-structural protein, NS: Non-structural protein (accessory protein).

Distribution of mutations characterizing VOCs and VOIs

According to WHO, SARS-CoV-2 is currently found as several grouped variants, into VOIs and VOCs variants, each defined by multiple mutations.

We found in this study, in addition to the S-D614G mutation common to the two variants VOI and VOC, associated mutations with VOCs: N501Y, the 70del (V70del) and H69del (H69del) at a low frequency (0.61 %) and the L452R mutation (2.08%).

We also observed, in addition to the D614G, mutations associated with VOIs: S-N439K, S-S477N at very low frequencies (<0.36%) and S-L452R (2.08%), which was also found in this study associated with VOC.

Eight proteins (NP1, NSP5, NSP7, NSP9, NSP11, ORF6, ORF10 and E) showed no mutable positions.

Distribution of mutable positions in proteins

Considering the FMPP, the mutable positions had the following distribution: NS7b-40 (42%)> NS7a-82 (15%) and NS7a-120 (15%)> M-80 (8%)> NSP6 -77 (6%), N-203 (6%)> NSP12-323 (5%)> NSP4-167 (4%), NSP4-492 (4%), NS3-26 (4%)> NS8-27 (3%), NS8-52 (3%), NS8-68 (3%) and NS8-73 (3%), NSP13-77 (3%), NSP14-394 (3%), S-614 (3 %)> NSP2-85 (1%), NSP3-488 (1%), NSP3-1228 (1%), NSP3-1469 (1%), NSP8-178 (1%), NSP10-104 (1%) , NSP15-127 (1%), NSP16-216 (1%) (Figure 3).

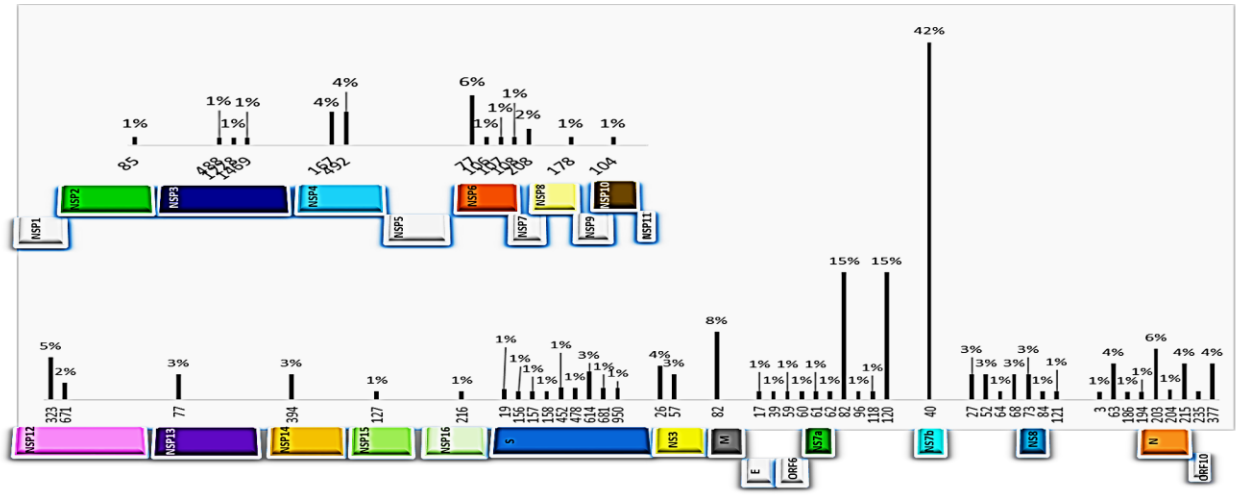

Figure 3: Percentage distribution of mutable positions by SARS-CoV-2 proteins.
S: Spike, N: Nucleocapecid protein, M: Membrane, NSP: Non-structural protein, NS: Non-structural protein (accessory protein).

However, when excluding the proteins without mutation (NSP7, NSP9, NSP11, E, ORF6 and ORF10), it appears that the only proteins which do not have mutable positions were NSP1 and NSP5.

Distribution of GISAID clades

The mutations found in the 44 sequences studied determined five (5) GISAID clades: G, GH, GR, GK and O (Figure 4). Considering only the periods when the virus sequences were available on the GISAID platform (March 2020 to July 2021), it can be seen that during the month of June 2020, three clades were present: GH and GR with a predominance of GR. The G clade was under represented. In February 2021, the clades present were G and O. The O clade represented 14% of the sequences during the period from February to March 2021. Then, in June 2021, the only clade that persisted was GK, which corresponds to the delta variant (B.1.617.2 according to the PONGO nomenclature).

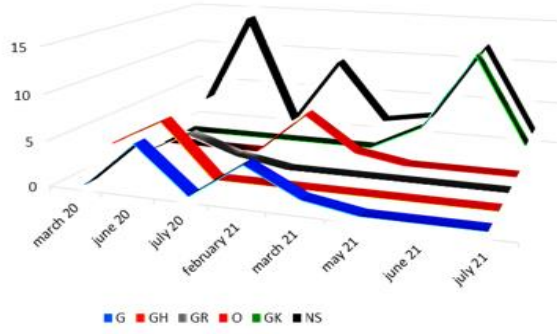

Figure 4: Temporal distribution of GISAID clades during the COVID-19 pandemic in Algeria (March 2020 to July 2021). NS: number of sequences; G, GH, GR, GK and O: GISAID clades.

Discussion

Distribution of mutations and their recurrences in virus genome

Unlike other studies, we found six proteins (NSP7, NSP9, NSP11, ORF6, ORF10 and E) among the 26 without mutation and 2 proteins without any mutable positions (NSP1 and NSP5).

In the mutated proteins, the S-D614G mutation was predominant (Figure 1). The replacement of D (aspartic acid) by G (glutamine) at position 614 of protein S is a highly recurring event in this study (97% of the sequences) and this despite the fact that position S-614 is one of the weakest protein's mutable virus (Figures 3 and 4). This mutation quickly became dominant (69.2% of sequences) in over 70 countries [7-8-9-10-4].

It is located in the Spike's S1 subunit, which is related to the host cell's ACE2 receptor-binding domain. The G acid at position 614 has been found to improve the Spike's affinity for the ACE2 by promoting greater exposure of the RBD (Receptor Binding Domain) of the virus [11].

In this study, the NSP12-P323L mutation is the second most frequent mutation (figure 2) and the most recurrent (97% of sequences), although it is found at a weakly mutable position (NSP12-323) (Figures 3 and 4). It is very common (69.47%) in 71 countries [12-13-14]. It sits in RdRp (RNA-dependent RNA polymerase), which is a key enzyme in the synthesis of viral RNA. It consists of the replacement of P (proline) by L (leucine) at position 323.

The S-D614G and NSP12-P323L mutations are concomitant, in our study, they are found simultaneously in 97% of the sequences (Figure 3). This observation has been reported in most populations. Indeed, from the start of the COVID-19 pandemic, these two mutations were found together at very high frequencies in the first sequences of SARS-CoV-2. In addition, as of January 2020, they are in the majority among hospitalized patients [4]. It appears that these two alterations in the SARS-CoV-2 genome are necessary for the epidemiological propagation of the virus.

When looking at the proteins of the viral RNA replication machinery (Nsp7 to Nsp16), we only find highly recurrent mutations: NSP12-P323L (97%), NSP13-P77L (40.91%) and NSP14-A394V (40.91%) (Figure 3). No moderately or weakly recurrent mutations are observed in the proteins of this machinery. This observation is also reported by the team of Ma Y and colleagues, showing the importance of these mutations in the multiplication of the virus [15]. They are a good target for anti-COVID-19 therapies [16].

Distribution of mutations characterizing VOCs and VOIs

In addition to the D614G mutation already described, we found four mutations associated with VOCs: the N501Y, L452R, V70del and H69del mutations.

The N501Y substitution was found in the VOCs variants Alpha, Gamma and Beta, first identified in the UK, Brazil and South Africa respectively. This mutation would increase transmissibility and affinity for the ACE2 receptor [17].

The L452R substitution is associated with Delta VOC and Iota/Kappa VOIs, first appearing in India, and significantly reduces the susceptibility of the virus to immunotherapy (https://www.fda.gov, updated November 17, 2021).

The V70del and H69del deletions are associated with the Alpha and Eta VOCs, first identified in the UK and UK/Nigeria respectively. These mutations are associated with increased infectivity and reduced neutralization of polyclonal serum from people cured of infection [18-19]. They are also associated with the S gene target failure (SGTF) during the multiplex RT-PCR test (https://www.finddx.org/covid-19/novel-variants/, updated January 7, 2021).

Regarding the mutations associated with the VOIs, we found, in addition to the D614G mutation, the substitutions S-S477N, S-N439K, S-L452R. The S-S477N substitution is associated with VOI Iota, which first appeared in the United States of America (New York). It would be responsible for the escape of the virus from neutralizing antibodies.

The N439K substitution associated with VOI B1 (PONGO nomenclature) was first identified in Scotland. It confers resistance to several neutralizing monoclonal antibodies and reduces the activity of polyclonal serum from people who have recovered from the infection. It is one of the immune escape mutations currently found in emerging SARS-CoV-2 variants [6].

Distribution of mutable positions in proteins

We also showed in this study that the most mutable positions were surprisingly found in accessory proteins NS7a/b (NS7a-82 and NS7b-40). Although these proteins are not essential neither for the replication nor for the infectivity of the virus, they would be, in particular NS7a, responsible for an opposition to the inflammatory reaction mediated by INF-1 during the COVID-19 by blocking of the STAT2 pathway [20].

When we accumulated the frequencies of mutations and mutable positions, only NSP1 and NSP5 proteins were found, which did not present any mutable positions (according to the criteria of this study). This observation could be useful in the development of molecular tests for the detection of SARS-CoV-2, which would be effective independently of the variants present. Indeed, the tests currently marketed for molecular diagnosis (known as PCR test) target regions of the genome often affected by mutations (S, N, RdRpand E), which could lead to false negatives [21]. It would be interesting to combine new PCR targets (NSPs genes) and classic targets (S, Nand E), in such a way as to detect all the variants of the virus.

Distribution of GISAID clades

The analysis of the distribution of the GISAID clades present during the period March 2020 to July 2021, showed that the G, GH, GR clades were present at the same time, then at the end of this period, the GK clade became predominant, as reported in several other works [22-23].

The GR and GH clades had a higher prevalence than the G clade, and then they were replaced by the GK clade. This clade is currently the most responsive. Overall, these observations have been verified in most countries [24-9]. The differences in the distribution of clades that can be observed between countries are due

to the unavailability of sequences or to delays between sampling and publication of results.

Conclusion

The virus circulating in Algeria during the period March 2020 to July 2021 presented 20 mutated proteins, of which S, N, NSP3 and NSP12 had the most mutations. The most frequent mutations were D614G and P323L found in the S and NSP12 proteins, respectively. These mutations are concomitant; they seem to be necessary for the epidemiological propagation of the original virus (Wuhan virus, China). Also of the 35 highly recurring mutations found in this study, D614G and P323L were the most frequent in the sequences examined.

In addition to the S-D614G mutation, we found four mutations associated with VOCs and three associated with VOIs. The S-L452R mutation is common in two variant types, it is the third most frequent after the S-D614G and NSP12-P323L mutations in the sequences studied.

When looking for the most mutable positions in proteins, we found that NS7a and NS7b (accessory proteins) are surprisingly the most mutable of all the proteins in the virus. These proteins are believed to be, in particular NS7a, involved in the anti-inflammatory effect by promoting immune escape.

Finally, the study of the GISAID clades carried out in this work showed that the G, GH and GR clades were present at the beginning of the March 2020 period, and then gradually they were replaced by GK, corresponding to the Delta variant (B.1.617.2) in July 2021, which is currently predominant in Algeria.

Conflicts of interest

No conflicts of interest.

Funding

This research did not receive any external funding

References

1. Kim D , Lee JY , Yang JS , Kim JW , Kim VN, Chang H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cellule*, 181(4): 914–921.e10.
2. Naqvi AAT, Fatima K, Mohammad T, Urooj Fatima, Indrakant K S, Archana S, Shaikh Muhammad A, Gururao H, Gulam M H, Md. Imtaiyaz H. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim Biophys Acta Mol Basis Dis*, 1: 1866(10), 165878.
3. Duffy S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biol,* 16(8): e3000003. https://doi.org/10.1371/journal.pbio.3000003.
4. Patro LPP, Sathyaseelan C, Uttamrao PP, Rathinavelan T. (2021). Global variation in SARS-CoV-2 proteome and its implication in pre-lockdown emergence and dissemination of 5 dominant SARS-CoV-2 clades. *Infect Genet Evol*, 93: 104973 published online June 18, 2021, doi: 10.1016/j.meegid.2021.104973, last access September 7, 2021.
5. Tian D, Sun YH, Zhou JM, Ye Q. (2021). The global epidemic of SARS-CoV-2 variants and their mutational immune escape. *J Med Virol*, published online August 18, 2021, doi: 10.1002/jmv.27376, last access October 3, 2021.
6. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JCC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife*, 9: e61312.
7. Daniloski Z, Guo X, Sanjana NE. (2021). The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. *eLife,* published online February 11, 2021, 10:e65365 doi: 10.7554/eLife.653657, last access October 10, 2021.
8. Klumpp-Thomas C, Kalish H, Hicks J, Mehalko J, Drew M, Memoli MJ. (2020). D614G spike variant does not alter IgG, IgM, or IgA spike seroassay performance. *medRxiv*. published online July 2020, doi: 10.1101/2020.07.08.20147371, last access September 4, 2021.
9. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W. (2020) Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 183: 812–827.
10. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T. (2020). The D614G

mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *BioRxiv*, 06.12.148726, published online Jun 12, 2020, doi:10.1101/2020.06.12.148726, last access September 2021.

11. Moore MJ, Dorfman T, Li W, Wong SK, Li Y, Kuhn JH, Coderre J, Vasilieva N, Han Z, Greenough TC, Farzan M, Choe H. (2004). Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. *J Virol*, 78(19):10628-10635.
12. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med*, 18(1):1-9.
13. Chand GB, Banerjee A, Azad GK. (2020). Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. *PeerJ*, 8: e9492, publisher online July 3, 2020. doi: 10.7717/peerj.9492, Last access September 6, 2021.
14. Zhao J, Zhai X, Zhou J. (2020). Snapshot of the evolution and mutation patterns of SARS-CoV-2. *bioRxiv*, published online July 5, 2020, doi: 10.1101/2020.07.04.187435), last access September 4, 2021.
15. Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U. S. A*, 112: 9436–9441.
16. Morse JS, Lalonde T, Xu S, Liu WR. (2020). Learning from the Past: Possible Urgent. Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *Chembiochem*, 21(5):730-738.
17. Luan B, Wang H, Huynh T. (2021). Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. FEBS Lett. *FEBS letters*, 595(10), 1454-1461.
18. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G. (2021). Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science*, 371 (6534):1139-1142.
19. Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S. (2021). Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host & Microbe*, 29:477–488. e4.
20. Cao Z, Xia H, Rajsbaum R, Xia X, Wang H, Shi PY. (2021). Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. *Cell Mol Immunol*, 18(3),746-748.
21. Artesi M, Bontems S, Göbbels P, Franckh M, Maes P, Boreux R, Meex C, Melin P, Hayette MP, Bours V, Durkin K. (2020). A recurrent mutation at position 26340 of SARS-CoV-2 is associated with failure of the E gene quantitative reverse transcription-PCR utilized in a commercial dual-target diagnostic assay. *Journal of clinical microbiology*, 58(10):e01598-20.
22. Bui NN, Lin YT, Huang SH, Lin CW. (2021). The extent of molecular variation in novel SARS-CoV-2 after the six-month global spread. *Infect Genet Evol*, 91:104800.
23. Hamed SM, Elkhatib WF, Khairalla AS, Noreddin AM. (2021). Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Sci Rep*, 11(1):1-8.
24. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole Á, Pereyaslov D, The WHO European Region sequencing laboratories and GISAID EpiCoV group. (2020). Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill*, 25(32):2001410.