



ROYAUME DU MAROC
Université Mohammed V - Rabat
Faculté de Médecine et de Pharmacie



RABAT

Année : 2022

N° : MM152022

MEMOIRE DE MASTER

MASTER DE **biotechnologie médicale**

OPTION : **biomédicale**

Intitulé :

ANALYSE GENOMIQUE DE SEQUENCES MAROCAINE DU COVID-19

Soutenu le 20/10/2022 par :

Salma CHAHED

Devant le jury composé de :

Pr LOUATI SARA

Faculté de Médecine et de Pharmacie Rabat

Président de jury

Pr AANIZ TARIK

Faculté de Médecine et de Pharmacie Rabat

Examineur

Pr OUADGHIRI MOUNA

Faculté de Médecine et de Pharmacie Rabat

Encadrant

Mme BENDANI HOUDA

Faculté de Médecine et de Pharmacie Rabat

Co-Encadrant

Dédicace

Je dédie ce travail :

A mes parents pour le soutien qu'ils m'ont apporté tout au long de mes études et aussi la confiance qu'ils ont eu en moi.

A mon mari Imad pour tous ses sacrifices, son amour, sa tendresse, son soutien tout au long de mes études.

A mon frère Ilyass et mes sœurs Hafsa et Nassma

A mes filles Loujayne et Julia

A tous ceux qui m'aiment

Remerciements

Au doyen de la faculté de médecine et de pharmacie de rabat pour m'avoir accueilli dans votre honorable établissement.

Au directeur du laboratoire de biotechnologie médicale à la faculté de médecine et de pharmacie de rabat professeur A. IBRAHIMI, de m'avoir accepté dans votre laboratoire en tant qu'étudiante.

A notre coordinatrice de l'option biotechnologie médicale Professeur OUADGHIRI MOUNA pour m'avoir fait confiance, merci d'accepter de m'avoir encadrer, merci pour votre écoute et votre accompagnement

A mon co-encadrante de Mémoire Mme Bendani Houda, j'aimerais chaleureusement la remercier, elle a accepté de m'accompagner dans des circonstances particulières, afin que je puisse enfin relever ce défi stimulant. Je tiens à vous exprimer ma profonde reconnaissance pour votre disponibilité, votre présence, votre patience et vos conseils.

Mes remerciements vont aussi à mes parent, mon mari, et mes petites filles, vous m'avez encouragé depuis le début jusqu'aujourd'hui malgré toutes les difficultés que j'ai rencontrées, Je vous remercie pour l'amour et le soutien que vous m'avez toujours apporté.

Résumé

Coronavirus (COVID19) est une maladie infectieuse due au virus SARS-CoV-2, découverte à Wuhan en décembre 2019, il s'agit d'une maladie infectieuse extrêmement contagieuse et très réactive. La première apparition de cas et de symptômes en dehors du Chine est en Thaïlande le 13 janvier 2020. Le premier cas importé a été détecté au Maroc le 02 Mars 2020, alors que le premier cas de transmission locale a été enregistré le 13 Mars 2020.

Le nombre de cas confirmés a augmenté peu à peu, et la propagation du virus s'est augmentée de plus en plus. Le virus a une nature changeante, ce qui lui a poussé à donner de nouvelles variantes plus dangereuses. En raison de cette évolution constante, il est très productif de disposer de séquences continues du génome et d'effectuer des études génomiques approfondies des différentes lignées disponibles.

La plupart des pays ont donc mis en place des réseaux de séquençage génomique pour aider à accomplir cette mission. Le Maroc, avec ces nombreux laboratoires de recherches, détecte et confirme la circulation des variants et leur propagation dans toutes les régions du royaume.

Dans ce cadre, nous avons effectué l'analyse génomique d'un ensemble de séquences marocaines du SARS-COV 2 appartenant à différentes périodes de la pandémie. Un ensemble de traitement a été effectué commençant par le contrôle qualité, jusqu'à l'analyse et l'identification des lignées en utilisant plusieurs outils et commandes. Nous avons pu identifier les mutations propres à chaque variant, responsables de la sévérité des variants, la propagation du virus, et la gravité de la maladie en affectant des gènes responsables de l'infection et la défense immunitaire.

Mots clés : SARS-COV-2, variants, analyse génomique, mutations, protéine spike.

Abstract

Coronavirus (COVID19) is an infectious disease due to SARS-CoV-2 virus, discovered in Wuhan in December 2019, it is an extremely contagious and highly reactive infectious disease. The first occurrence of cases and symptoms outside of China was in Thailand on January 13, 2020. The first imported case was detected in Morocco on 02 March 2020, while the first locally transmitted case was recorded on 13 March 2020.

The number of confirmed cases has been increasing gradually, and the spread of the virus has been increasing more and more. The virus has a changing nature, which has caused it to give new and more dangerous variants. Because of this constant evolution, it is very productive to have continuous genome sequences and to perform thorough genomic studies of the different lineages available.

Most countries have therefore set up genomic sequencing networks to help accomplish this mission. Morocco, with its numerous research laboratories, detects and confirms the circulation of variants and their spread in all regions of the kingdom.

Within this framework, we performed genomic analysis of a set of Moroccan SARS-COV 2 sequences belonging to different periods of the pandemic. A set of processing was performed starting with quality control, up to lineage analysis and identification using several tools and commands. We were able to identify the mutations specific to each variant, responsible for the severity of the variants, the spread of the virus, and the severity of the disease by affecting genes responsible for infection and immune defense.

Key words: SARS-COV-2, variants, genomic analysis, mutations, spike protein.

ملخص

فيروس كورونا (كوفيد 19) هو مرض معد بسبب فيروس سارس-كوف-2 ، اكتشف في ووهان في ديسمبر 2019 ، وهو مرض معد للغاية وعالي التفاعل. أول ظهور للحالات والأعراض خارج الصين في تايلاند في 13 يناير 2020. تم اكتشاف أول حالة مستوردة في المغرب في 02 مارس 2020 ، بينما تم تسجيل أول حالة انتقال محلي في 13 مارس 2020

ازداد عدد الحالات المؤكدة تدريجيا ، وزاد انتشار الفيروس أكثر فأكثر. الفيروس له طبيعة متغيرة ، مما دفعه إلى إعطاء متغيرات جديدة أكثر خطورة. بسبب هذا التطور المستمر ، من المفيد جدا الحصول على تسلسلات مستمرة للجينوم وإجراء دراسات جينومية متعمقة لمختلف الخطوط المتاحة

لذلك أنشأت معظم البلدان شبكات التسلسل الجيني للمساعدة في إنجاز هذه المهمة. المغرب ، بمختبراته البحثية العديدة ، يكتشف ويؤكد تداول المتغيرات وانتشارها في جميع مناطق المملكة

في هذا السياق ، أجرينا التحليل الجينومي لمجموعة من التسلسلات المغربية لفيروس سارس-كوف 2 التي تنتمي إلى فترات مختلفة من الوباء. تم تنفيذ مجموعة من المعالجة بدءا من مراقبة الجودة ، وصولا إلى تحليل وتحديد الخطوط باستخدام العديد من الأدوات والضوابط. تمكنا من تحديد الطفرات الخاصة بكل متغير ، المسؤولة عن شدة المتغيرات ، وانتشار الفيروس ، وشدة المرض من خلال التأثير على الجينات المسؤولة عن العدوى والدفاع المناعي

الكلمات المفتاحية: سارس-كوف-2 ، المتغيرات ، التحليل الجينومي ، الطفرات ، بروتين سبايك

Liste des figures

Figure 1: Schéma de la structure du SARS-CoV-2.	5
Figure 2: Structure des coronavirus.	6
Figure 3: Représentation schématique du génome de SARS-CoV-2.	7
Figure 4: Classification et taxonomie des coronavirus humains (HCoV).	8
Figure 5: Schéma du cycle d'un coronavirus entrant dans une cellule et se répliquant à l'intérieur	10
Figure 6: Schéma des différents modes de transmission du SARS-CoV-2	12
Figure 7: Nombre de cas et de morts au monde.....	13
Figure 8: Nombre de cas et de morts au Maroc.....	14
Figure 9: Capture de l'interface de plateforme nexclade.....	20
Figure 10: Interface de la plateforme Pangolin.....	24
Figure 11: Contrôle qualité de la séquence Alpha.....	25
Figure 12: La page de résultats d'identification des lignées sur le site web Pangolin.....	26
Figure 13: Distribution des mutations de la séquence Alpha.	27
Figure 14: Distribution des mutations par genes pour la séquence DELTA.	30
Figure 15: Représentation des mutations par gène pour la séquence Mu.....	33
Figure 16: Représentation des mutations par gène de la séquence Omicron.....	36

Liste des tableaux

Tableau I: résultat d'identification des lignées des quatre séquences.....	26
Tableau II: : les délétions extraites de la séquence Alpha.	28
Tableau III: les SNP non synonymes codantes extraites de la séquence Alpha.	29
Tableau IV: les délétions extraites de la séquence Delta.	31
Tableau V: les SNP non synonymes codantes extraites de la séquence Delta.	32
Tableau VI: Les insertions extraites de la séquence Mu.....	34
Tableau VII: Les SNP non synonymes codantes extraites de la séquence Mu	35
Tableau VIII: les délétions extraites de la séquence Omicron.....	37
Tableau IX: Les SNP non synonymes extraites de la séquence Omicron	38

Liste des abréviations

ACE2 : Angiotensin-Converting Enzyme 2 (Enzyme de conversion de l'angiotensine 2)

ADNc : Acide désoxyribonucléique complémentaire

ARN : Acide ribonucléique

BAM : Binary Alignment Map (Carte d'alignement binaire)

BCF : binary call variant

CNRST : Centre National pour la Recherche Scientifique et Technique

COVID-19: Coronavirus disease 2019

COVs: Corona viruses

CQ : contrôle qualité

ECA : Enzyme de conversion de l'angiotensine 2

GISAID : Global Initiative on Sharing All Influenza Data

NGS : New generation sequencing

NSP : Protéines non structurales

OMS : L'organisation mondiale de la sante

ORF : Open reading frames

PCR : Polymerase chain reaction

S : la glycoprotéine Spike

SAM: Sequence Alignment Map (Carte d'alignement des séquences)

SARS : Le syndrome respiratoire aigu sévère

SARS-COV-2 : Le syndrome respiratoire aigu sévère de coronavirus 2

VCF : variant call format

Table des matières

Dédicace.....	2
Remerciements.....	3
Résumé	4
Abstract.....	5
ملخص.....	6
Liste des figures	7
Liste des tableaux.....	8
Liste des abréviations.....	9
Introduction	1
Revue bibliographique	3
I. Revue bibliographique	4
1. Covid 19.....	4
2. Caractéristiques morphologiques du SARS-CoV-2.....	4
3. Structure du génome	6
3.1 Structure virale.....	6
3.2 Structure génomique	7
4. Taxonomie.....	8
5. Cycle de vie	9
6. Transmission	11
7. Epidémiologie mondiale	12
8. Épidémiologie au Maroc	13
9. Surveillance génomique au Maroc	15
Partie pratique	16
II. Matériel et méthodes	17
1. Collection des données.....	17
1.1 Le séquençage ION torrent	17
1.2 GISAID	18
2. Analyse génomique.....	19
2.1 Contrôle de qualité	19
2.2 Indexation de la référence	21

2.3	Alignement de la séquence avec la référence	21
2.4	Convertir SAM to BAM, trier et indexer le fichier BAM	22
2.5	Variant calling	23
2.6	Annotation	23
3.	Analyse et identification des lignées.....	24
III.	Résultats.....	25
1.	Contrôle de qualité	25
2.	Identification des lignées	26
3.	Mutation des séquences :.....	27
3.1	Alpha :	27
3.2	Delta :.....	30
3.3	Mu :.....	33
3.4	Omicron :.....	36
IV.	Discussion.....	39
V.	Conclusion.....	41
VI.	Référence	42

Introduction

En Mars 2020, l'organisation mondiale de la santé (OMS) a déclaré, l'épidémie éclatée à Wuhan (Chine), comme une pandémie en raison de sa propagation rapide dans le monde (1)(2). La transmission par aérosols, surtout à proximité, est bien reconnue et semble être le mode prédominant de transmission du SARS-CoV-2. Le risque de transmission est augmenté dans des espaces restreints, mal ventilés ou à forte densité d'occupants (3)(4). Les symptômes causés par ce virus sont comparables à ceux de la grippe mais dans les cas les plus graves, l'infection peut provoquer une pneumonie ou des difficultés respiratoires (5). Lors de l'émergence de l'épidémie, l'une des tâches de santé publique étaient d'empêcher la propagation du virus. Le virus évolue avec le temps, il a connu des mutations ce qui a donné plusieurs variantes. Les erreurs de l'ARN polymérase virale durant la réplication sont responsables du caractère mutationnel du SARS-CoV-2, permettant ainsi l'émergence de nouvelles variantes génétiques (6).

Au Maroc, le premier cas enregistré était le 02 Mars 2020 qui était importé de l'étranger, puis le premier cas local par transmission a été détecté le 13 Mars 2020. Le nombre de cas confirmés a augmenté progressivement, amenant notre pays à mettre en œuvre des mesures de distanciation sociale depuis le 15 mars 2020, l'arrêt des études et des prières à partir du 16 mars 2020 et le confinement progressif depuis le 20 mars 2020. Ces mesures ont probablement permis un ralentissement relatif de la propagation de l'épidémie. Mais depuis la fin du confinement le pays a connu des vagues avec une élévation de nombre de cas infectés, de nombre de morts de nombres de patients hospitalisés en réanimation et en soins intensifs.

Les vagues que le Maroc a connu sont Alpha, Delta, Omicron BA1, et Omicron BA5(7).

Grâce aux techniques du séquençage à haut débit, la séquence de références de Wuhan est disponible depuis janvier 2020 dans la GenBank du National Center for Biotechnology Information (NCBI). Une analyse bio-informatique est alors nécessaire pour pouvoir caractériser le virus, mieux comprendre ses caractéristiques biologiques et cliniques et cartographier sa propagation.

L'objectif principal de cette étude est de comprendre l'épidémiologie du COVID au Maroc en analysant des séquences marocaines de différentes périodes de la pandémie,. Le but est d'identifier

les mutations les plus courantes et de comparer les variants, en passant par plusieurs étapes commençant par le contrôle qualité, analyse génomique, et identification des lignées.

Revue bibliographique

I. Revue bibliographique

1. Covid 19

Le Covid 19, maladie provoquée par le nouveau Coronavirus, découverte la première fois à Wuhan en décembre 2019, il s'agit d'une maladie infectieuse extrêmement contagieuse et très réactive, elle s'est propagée rapidement d'un pays à autre, l'organisation mondiale de la santé a déclaré la pandémie comme étant une urgence de santé publique d'intérêt international le 31 janvier 2020 après sa propagation rapide (8).

Les patients infectés par ce virus présentent une maladie respiratoire d'intensité légère à modérée, certains d'entre eux peuvent être rétablis sans avoir recours à un traitement particulier et certains nécessitent parfois une hospitalisation au service de soins intensifs ou en réanimation. Les sujets âgés, les patients immunodéprimés, les femmes enceintes, les personnes cardiaques, diabétiques, hypertendues, ou qui présentent une maladie respiratoire chronique, ont plus de risques de présenter une forme grave (9).

La transmission du virus SARS-CoV-2 se fait par des gouttelettes respiratoire (toux, éternuement, chant, crie). Il peut également se propager à travers les surfaces ou le touché d'objet contaminé par le virus, ou les personnes touchent ces derniers et se transmet au corps par la bouche, le nez, ou les yeux, surtout que la plupart ne se lavent par fréquemment les mains(10) .

La meilleure façon de se protéger contre l'infection est de maintenir la distance de sécurité avec tout le monde, porter le masque, laver fréquemment les mains, éviter les rassemblements, et tenir à jour la vaccination contre le Covid-19 dès que possible (11) .

2. Caractéristiques morphologiques du SARS-CoV-2

La pandémie actuelle de Covid-19, est causé par Le virus SARS-CoV-2 qui fait partie de la famille des coronavirus, ce sont des virus enveloppés possédant un génome à ARN de sens positif et avec une nucléocapside de symétrie hélicoïdal (12) .

Le noyau est entouré d'une enveloppe lipoprotéique qui se forme pendant le bourgeonnement du virus à partir des membranes intracellulaires, ils sont constitués de :

- La protéine Spike (S)(12) qui est une protéine de l'enveloppe, elle est présente dans tous les coronavirus, elle permet au virus de pénétrer dans les cellules.
- La protéine M (La glycoprotéine de la matrice), parce qu'elle traverse trois fois la bicouche lipidique, est considérée comme un composant de la structure interne du noyau et de l'enveloppe virale, elle joue un rôle important dans la formation intracellulaire des particules virales.
- La protéine d'enveloppe (E) fait également partie de l'enveloppe virale, mais en quantité beaucoup plus faible que les autres protéines d'enveloppe virale.

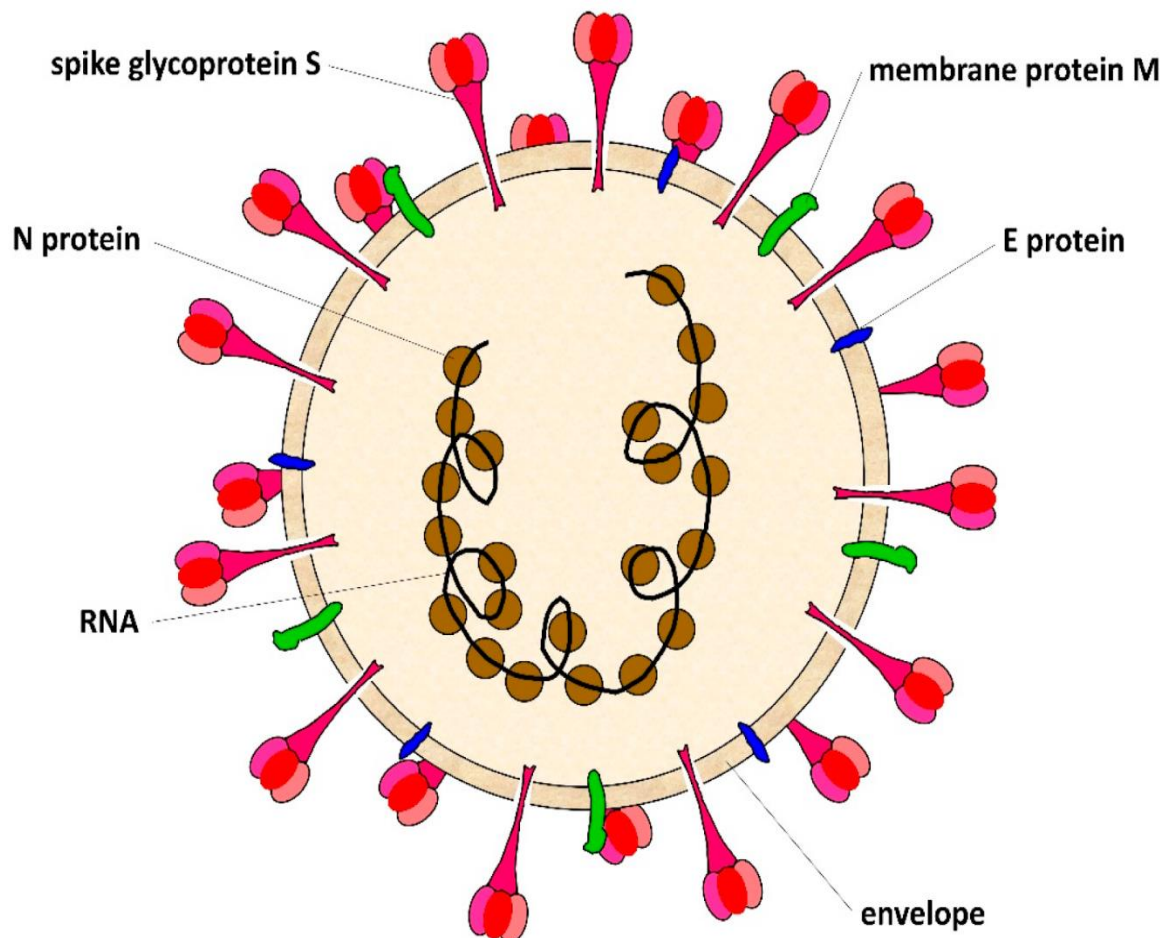


Figure 1: Schéma montre la structure du SRAS-CoV-2. (13)

3. Structure du génome

3.1 Structure virale

Le Virus est sphérique, d'une enveloppe de 60 à 220 nm, comprend de l'extérieur, vers l'intérieur, la glycoprotéine (S) Spike (donne l'aspect d'une couronne au virus en microscopie électronique), l'enveloppe, la membrane et la nucléocapside elle-même, icosaédrique à symétrie cubique. Cette dernière contient une molécule de génome viral : de l'acide ribonucléique (ARN) monocaténaire, non segmenté et positif (29 881 paires de bases) (14).

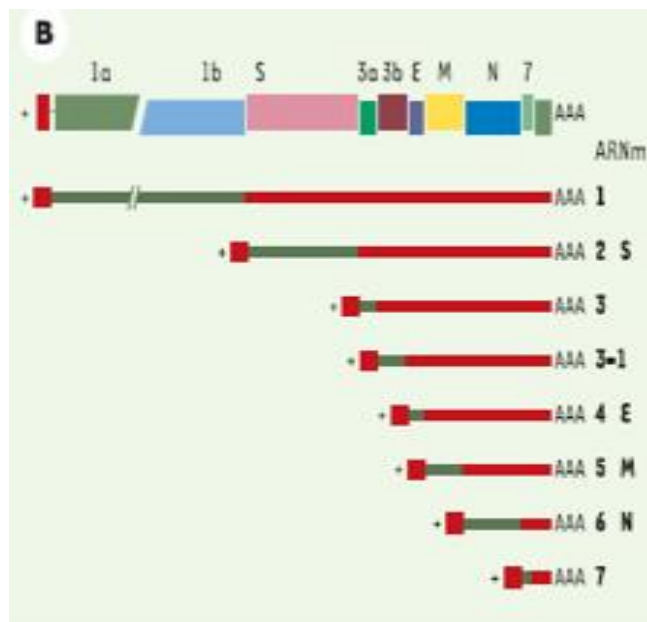


Figure 2: Structure des coronavirus.

Figure 2. Structure schématique de l'ARN génomique et des ARN subgénomiques d'un coronavirus prototype. L'ARN génomique (brin +) code pour les protéines d'enveloppe et de nucléocapside ainsi que pour la réplicase, transcrite à partir de l'ORF (open reading frame) 1a puis de l'ORF 1b par changement de phase de lecture. La polyprotéine produite par l'ORF 1a/1b est ensuite protéolysée en diverses protéines qui forment le complexe réplcatif. Les protéines structurales (S M N et E) sont traduites à la première phase de lecture (en vert) des ARNm initiées en aval dans la séquence génomique du virus. À l'extrémité 5' des ARN, une séquence 5'-leader est présente, identique à l'extrémité 5' de l'ARN génomique (boîte rouge). AAA = polyadénylation (d'après (15))

3.2 Structure génomique

Le génome du virus est constitué d'un ARN linéaire simple brin, non segmenté, de polarité positive (qui peut donc être directement traduit en protéine), d'environ 30 kb (le plus grand des virus à ARN) (16) et qui code pour 7 à 10 protéines (17).

Le génome comporte un nombre divers de cadres de lecture ouverts (*ORF*). Le premier ORF (ORF1a/b) comporte les deux tiers de l'ARN viral, il traduit 2 polyprotéines (pp1a et pp1b), et génère seize protéines non structurales (SNP), tandis que les autres ORF restants codent pour des protéines accessoires et de structure.

Quatre protéines essentielles de structure sont codées par le reste du génome du virus, dont la glycoprotéine (S), la protéine de l'enveloppe (E), la protéine matricielle (M) et la protéine nucléocapside (N), ainsi que plusieurs protéines accessoires, qui interfèrent avec la réponse immunitaire de l'hôte (18).

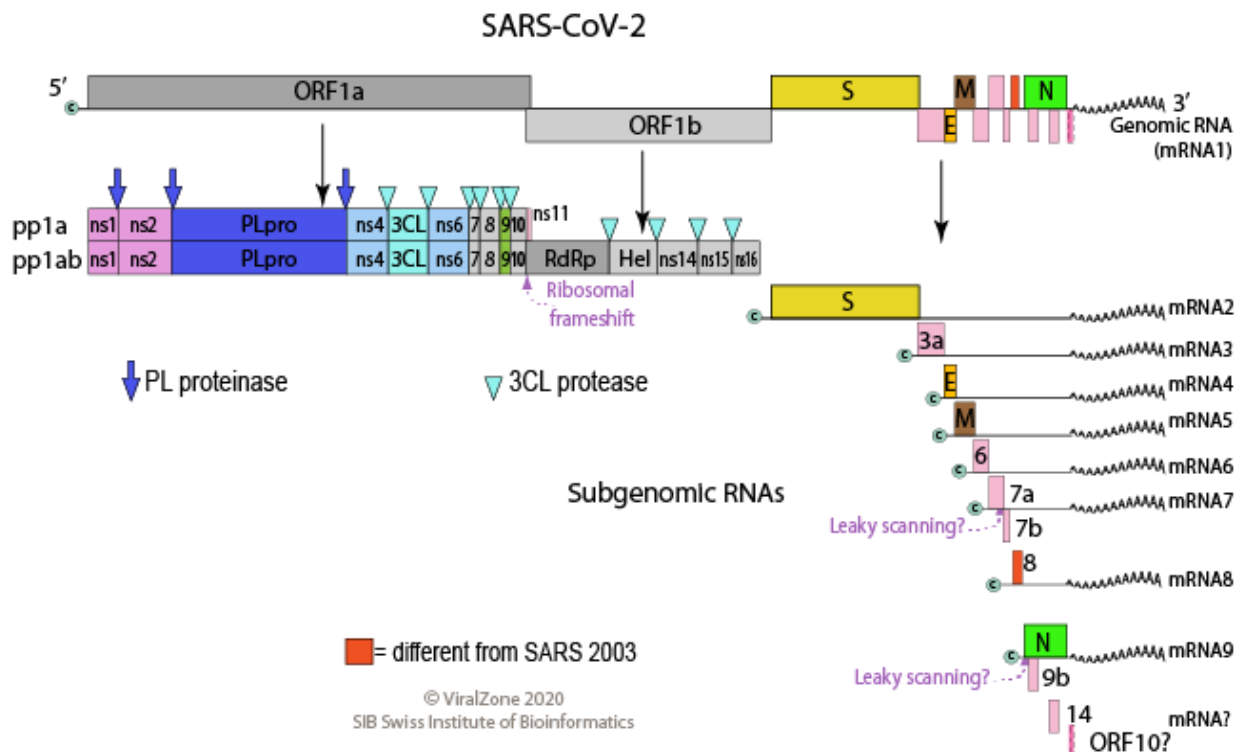


Figure 3: Représentation schématique du génome de SRAS-CoV-2.

4. Taxonomie

Les coronavirus constituent le plus grand groupe de virus appartenant à l'ordre des nidovirales, C'est l'ordre qui comprend les familles coronaviridae, arteriviridae, mesoniviridae et roniviridae, Les Coronaviridae sont subdivisés en quatre genres : (les coronavirus alpha, bêta, gamma et delta), Les anciens CoV de types 1, 2 et 3 sont remplacées par les Alpha-, Bêta- et Gamma Coronavirus. En 2011, le genre Deltacoronavirus a été défini et regroupe principalement des virus aviaires. Le genre Betacoronavirus est segmenté en 4 clades « a, b, c et d » ou « A, B, C, et D » (Figure 4). Les coronavirus (CoVs) sont responsables d'infections respiratoires et digestives chez plusieurs oiseaux et mammifères, six étaient connus comme responsables d'infections humaines : deux alphacoronavirus, et quatre betacoronavirus (19).

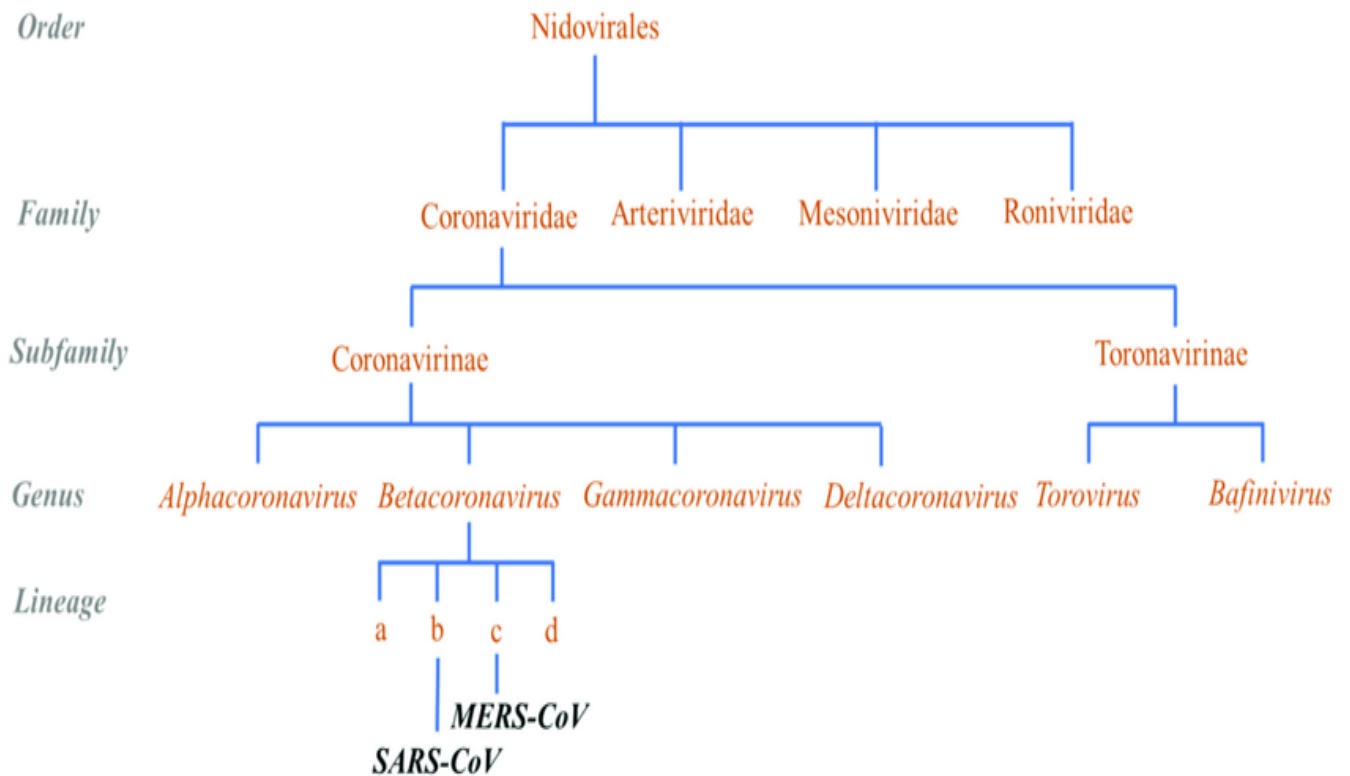


Figure 4: Classification et taxonomie des coronavirus humains (HCoV).

5. Cycle de vie

Le cycle de vie répliatif des virus varie fortement en fonction de l'espèce et de la catégorie du virus. L'un des principaux déterminants de l'initiation et de la progression de l'infection à SARS CoV-2 est l'entrée du virus et l'infection des cellules de l'hôte (20) .

La Figure 5 montre les différentes étapes dans le cycle de répliation de SARS COV-2. Ce cycle en principe possède quatre étapes principales, qui commencent tout d'abord avec :

- L'attachement de la protéine réceptrice, la protéine spike à l'enzyme de conversion propre à l'angiotensine (ACE2) (21). Cet attachement est fait par une signalisation cellulaire entre ECA2 et la protéine S.
- Pénétration intracellulaire du virus,
- L'expression de l'enzyme répliacase,
- Répliation et la transcription de l'ARN viral et l'assemblage et la libération des virions.

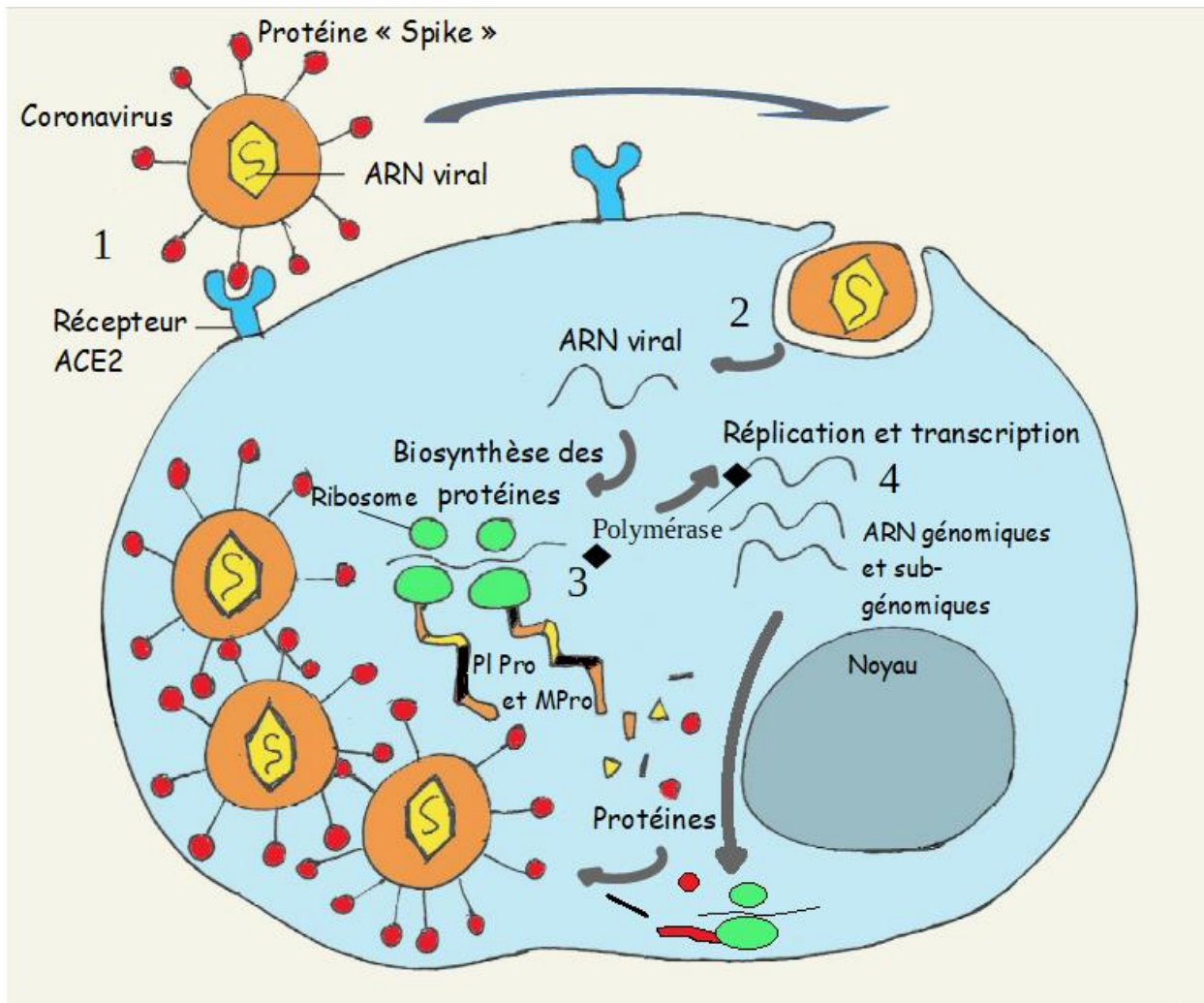


Figure 5: Schéma du cycle d'un coronavirus entrant dans une cellule et se répliquant à l'intérieur [8].

Le virus se lie grâce à ses protéines de surface, les protéines « Spike », à un récepteur membranaire de la cellule pour entrer dans la cellule. Pour le SARS-CoV-2, ce récepteur est l'enzyme de conversion de l'angiotensine 2 : ACE2. Ce récepteur est présent sur les cellules de différents organes : le nez, les yeux, les poumons, le système digestif, le cœur et dans une moindre mesure dans les reins et le foie. Le SARS-CoV-2 peut donc infecter tous ces organes. **(1)**

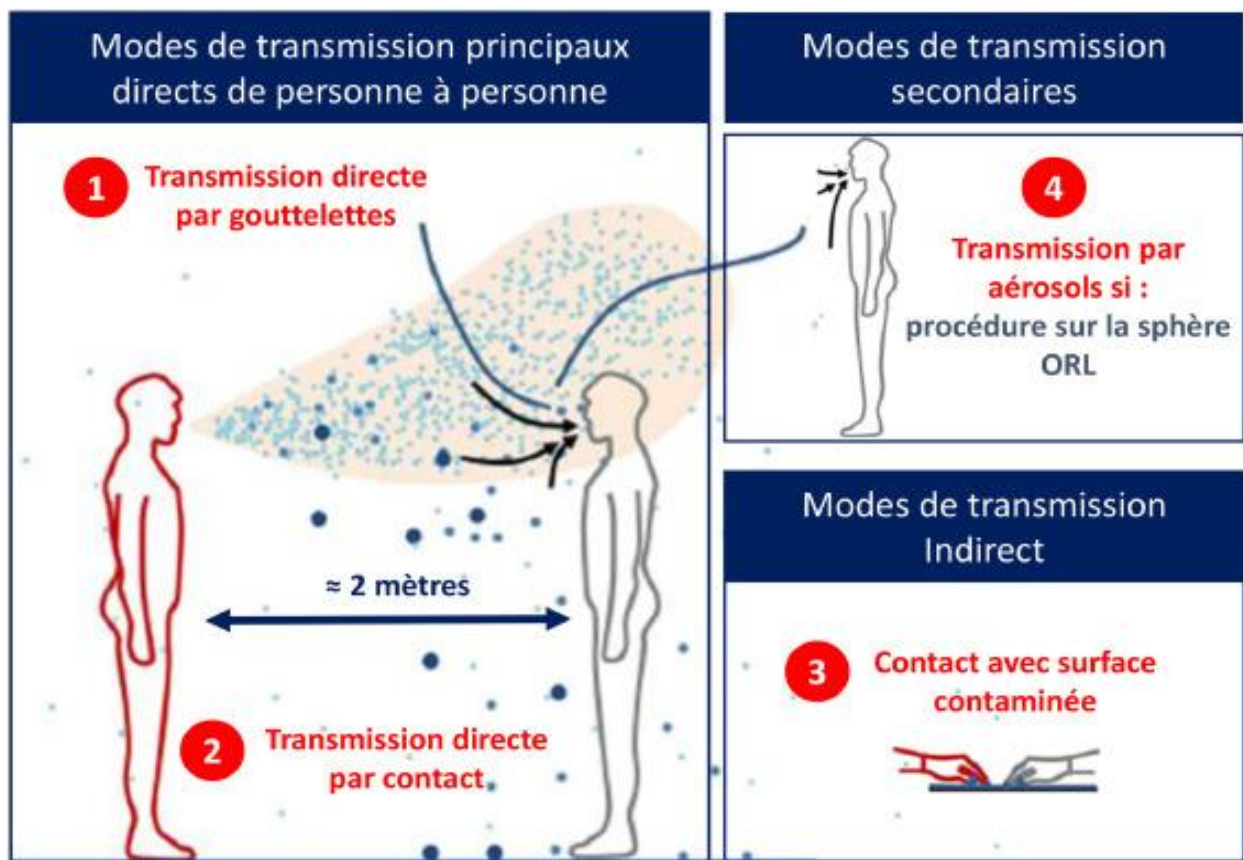
Le génome du coronavirus est un ARN simple brin, il est libéré lors de l'entrée virale ; cela démarre le processus de traduction de l'ARN génomique viral au niveau des ribosomes qui conduit à la synthèse de deux grandes protéines **(2)**, coupées ensuite par 2 protéases : une protéinase (appelée Mpro) et la protéase de type papaine (PI pro) ce qui donne des protéines non structurales comme la polymérase **(3)** (22).

Cette polymérase réalise ensuite une réplication de l'ARN génomique du virus et une transcription de cet ARN ce qui permet de former des ARN messagers appelés ARN « subgénomiques » (4). La traduction de ces ARN « subgénomiques » par les ribosomes permet la formation des protéines structurales du virus (5).

Ces protéines seront assemblées avec l'ARN génomique viral pour former les nouveaux virus (6). Les étapes 3 et 5 nécessitent la machinerie cellulaire sans laquelle les protéines virales ne pourraient être produites.

6. Transmission

La transmission des infections respiratoires se fait en plusieurs étapes successives. Tout d'abord, le virus est présent au niveau d'un réservoir, d'un individu infecté. Ce dernier doit parler cracher tousser ou éternuer, ces gestes permettent le jet des sécrétions respiratoire dans l'air, de différentes tailles sous formes de gouttelettes(23). Les plus grosses sédimentent immédiatement après émission. Les plus petites gouttelettes forment des droplet nuclei ces derniers restent dans l'air. En cas d'infection respiratoire, ces sécrétions peuvent comporter des micro-organismes, le mode de transmission au récepteur selon le micro-organisme peut être soit par l'inhalation des droplet nuclei, avec les grosses gouttelettes, par contact des muqueuse (nasal, conjonctivale, buccale). De manière indirecte cette fois, les mains du sujet réceptif peuvent se contaminer au contact de sécrétions ORL du patient ou d'une surface, puis être portées au visage (24).



- Larges gouttelettes
- Droplet nuclei

Figure 6: ce Schéma montre les différents modes de transmission du SARS-CoV-2 (24).

7. Épidémiologie mondiale

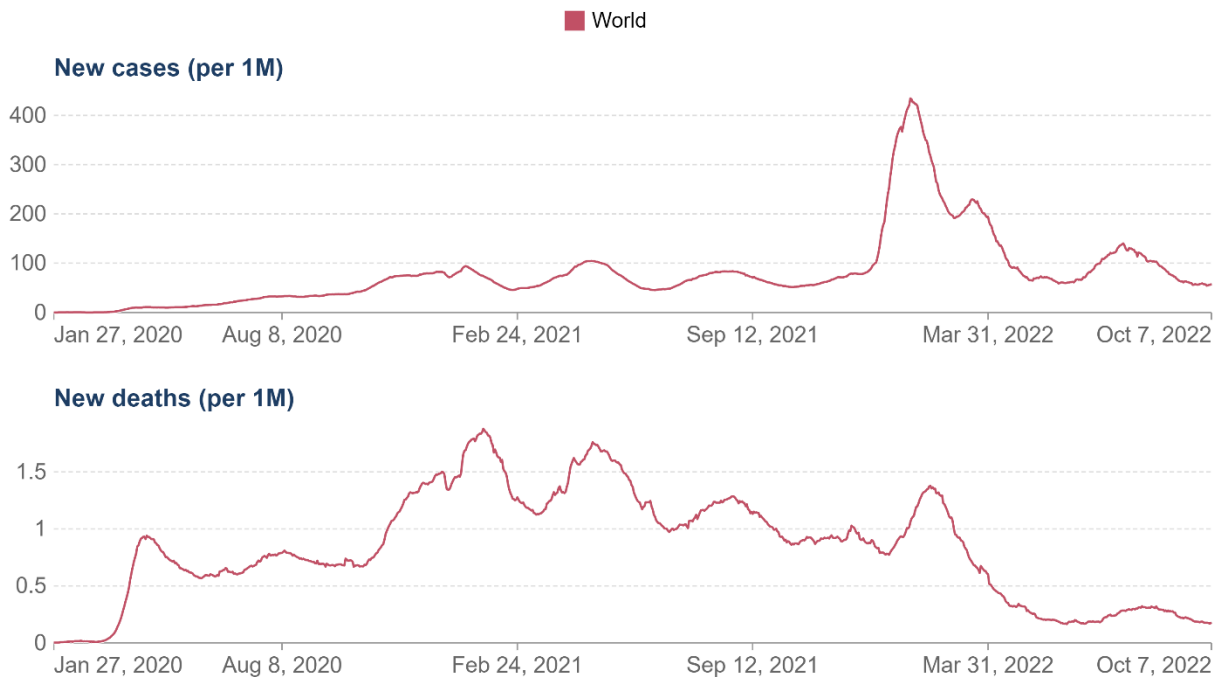
Les autorités sanitaires de Wuhan, le 31 décembre 2019, ont signalé un groupe de cas infectés de pneumonies de cause inconnue. La grande majorité des cas avaient visité le marché alimentaire qui vend (des poissons, des fruits de mer et des animaux vivants). Ce marché a été fermé le 1 janvier 2020. L'investigation a montré que le premier groupe de personnes infectées concernait 41 patients qui ont développé des symptômes entre le 8 décembre 2019 et le 2 janvier 2020. Le virus s'est rapidement diffusé en dehors des frontières chinoises (25).

En janvier 2020, les premiers malades ont été diagnostiqués en Europe, notamment en France. L'Italie a connu la première vague de contamination un mois plus tard.

Le nombre de morts s'est progressivement réduit suite aux restrictions sanitaires mises en place combinées aux campagnes vaccinales dans différents pays.

Daily new confirmed COVID-19 cases & deaths per million people

7-day rolling average. Limited testing and challenges in the attribution of cause of death means the cases and deaths counts may not be accurate.



Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

Figure 7: Nombre de cas, de morts au monde.

8. Épidémiologie au Maroc

Devant l'évolution rapide de la situation épidémiologique internationale du COVID-19, l'Organisation Mondiale de la Santé l'a décrété « Urgence de Santé Publique, de Portée internationale » et cela le 30 janvier 2020, puis pandémie le 12 mars 2020.

La chine est la source de l'épidémie, c'est le pays qui a connu la plus grande épidémie de covid-19.

Le premier cas enregistré au Maroc le 02 Mars 2020 est importé de l'étranger, alors que le premier local par transmission a été détecté le 13 Mars 2020. Le nombre de cas confirmés a augmenté progressivement, amenant notre pays à mettre en œuvre des mesures de distanciation sociale, consistant en :

1) la fermeture des frontières terrestres, aériennes et maritimes depuis le 15 mars 2020 ;

2) l'arrêt des études pour tous les niveaux scolaires et universitaires à partir du 16 mars 2020

3) l'arrêt des prières au niveau des mosquées depuis le 16 mars 2020 ;

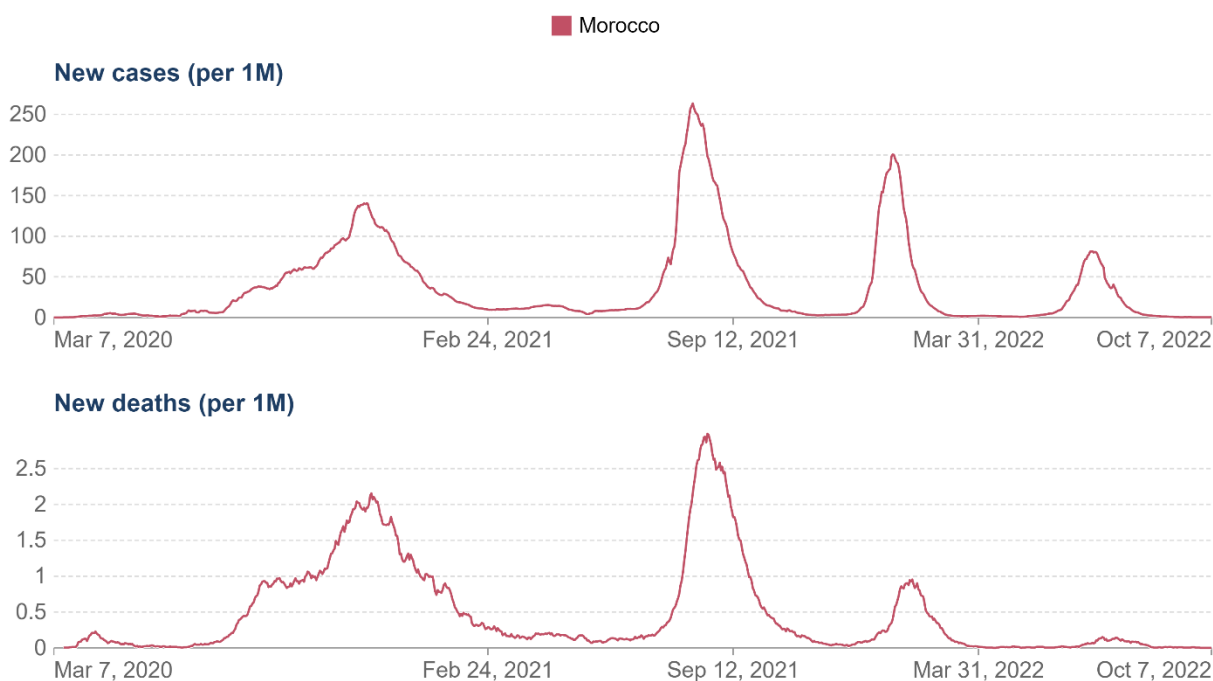
4) le confinement progressif de la population depuis le 20 mars 2020,

Ces mesures, dont l'impact doit être observé dans les 10 à 14 jours suivant leur entrée en vigueur, ont probablement permis un ralentissement relatif de la propagation de l'épidémie.

Daily new confirmed COVID-19 cases & deaths per million people

7-day rolling average. Limited testing and challenges in the attribution of cause of death means the cases and deaths counts may not be accurate.

Our World
in Data



Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

Figure 8: Nombre de cas, de morts au Maroc.

La première apparition de cas et de symptômes en dehors du chine est en Thaïlande le 13 janvier 2020, cela est après deux semaines de la première apparition (début de l'épidémie), le virus est transporté par une voyageuse chinoise (26).

Les pays les plus touchés depuis le début jusqu'à aujourd'hui sont : La Chine, l'Italie, l'Espagne, et les états unis.

Le Maroc a connu quatre vagues de variantes COVID 19, la première vague est celle de la variante Alpha, la deuxième du variant Delta, la troisième du variant Omicron BA1, et la dernière du variant Omicron BA5.

9. Surveillance génomique au Maroc

En raison de l'évolution constante du virus, il est très productif de disposer de séquences continues du génome. La majorité des pays ont mis en place, des réseaux de séquençage génomique, pour aider à accomplir cette mission. Le séquençage génomique permet de suivre les différentes mutations du virus et fournit ainsi des informations sur le statut du virus en termes de transmissibilité et de virulence, Au Maroc, le système de surveillance génomique du SARS-CoV2 est mis en place par le ministère de la Santé pour permettre la détection et la confirmation de la circulation des variants et de sa propagation dans toutes les régions du royaume. Plusieurs laboratoires s'associent pour réaliser un séquençage génomique régulier du SARS COV 2 sur l'ensemble du territoire marocain. Parmi eux l'Institut National d'Hygiène ; le Laboratoire de Biotechnologie Médicale (Med Biotech) de La Faculté de Médecine et Pharmacie Rabat ; Centre National de la Recherche Scientifique et Technique (CNRST), Laboratoire pasteur, et le Laboratoire de Recherche et d'Analyses Médicales de la Fraternelle de la Gendarmerie Royale. Le but est de comparer les séquences extraites par les échantillons avec les séquences de référence afin de détecter les mutations.

Partie pratique

II. Matériel et méthodes

1. Collection des données

L'identification de l'agent causal d'une infection inconnue passe généralement par le séquençage de son matériel génétique, c'est-à-dire la détermination de la séquence des nucléotides (les A, T, C et G) qui composent son génome, c'est le cas du virus SARS-COV-2.

Notre dataset se compose de 14 séquences marocaines collectées depuis GISAID et au sein du laboratoire.

- Deux séquences collecté depuis Gisaid.
- Douze séquences au laboratoire.

1.1 Le séquençage ION torrent

Le séquençage Ion Torrent est basée sur la détection d'ions hydrogènes libérés lors de la polymérisation de l'ADN. C'est une méthode de séquençage par synthèse durant laquelle un brin complémentaire est créé à partir d'une matrice. Le séquençage Ion Torrent est subdivisé à quatre étapes principales : (27)

- La construction de la banque.
- L'amplification.
- Le séquençage.
- L'analyse (28).

Contrairement au séquençage Illumina, le séquençage Ion Torrent ou Ion proton n'utilise pas de signaux optiques. Il exploite plutôt le fait que l'ajout d'un dNTP à un polymère d'ADN libère un ion H⁺. Comme dans d'autres types de NGS, l'ADN ou l'ARN d'entrée est fragmenté, en fragments d'environ 200 bp. Des adaptateurs sont ajoutés et une molécule est placée sur une bille. Les molécules sont amplifiées sur la bille par PCR en émulsion. Chaque bille est ensuite placée dans un seul puits d'une lame(29,30).

1.2 GISAID

GISAID est une initiative scientifique mondiale créée en 2008, elle a pour but de donner un accès libre aux données génomiques des virus de la grippe et du coronavirus responsable de la pandémie de COVID19, c'est à GISAID qu'étaient mises les premières séquences du génome entier du virus SARS COV 2, cela a permis de détecter le virus, de développer le vaccin, de faire des tests diagnostiques, et de bien comprendre la pandémie.

GISAID facilite l'épidémiologie génomique et la surveillance en temps réel de l'émergence de nouvelles souches virales COVID-19 à travers le monde, car sa base de données est la plus grande référence au monde en terme séquences du SARS-CoV-2.

À chaque découverte d'une nouvelle souche émergente de SARS-CoV-2 par le GISAID, avec une mutation de séquence potentiellement significative est signalée, les laboratoires seront invités à chercher et effectuer des tests sur des stocks cultivés en laboratoire afin de déterminer si les anticorps sont toujours capables de répondre efficacement à la nouvelle souche par rapport aux souches actuelles du SARS-COV-2, ces tests sont appelés test de neutralisation.

Si les tests effectués détectent des changements dans la capacité des anticorps à neutraliser la nouvelle souche virale SARS-COV-2, d'autres tests doivent être effectués afin d'évaluer s'il y a des changements dans la nouvelle souche, surtout en termes de gravité de la maladie, et la gravité de l'infection. Les résultats seront immédiatement partagés avec la communauté mondiale de la recherche via les réseaux de l'organisation mondiale de la santé, ce partage aide au développement du vaccin, et à déterminer si les vaccins existants protègent contre cette nouvelle souche. (31)

2. Analyse génomique

Dans cette étude, nous avons utilisé soit des commandes (**Ubuntu/ terminal**) ou des serveurs en ligne/application.

2.1 Contrôle de qualité

Pour cette étape, nous avons utilisé un serveur en ligne <https://clades.nextstrain.org/>. Il suffit de lui fournir notre séquence sous format FASTA et il nous retourne un ensemble d'informations y compris « Quality control ». Il vérifie la qualité des séquences en faisant plusieurs contrôles de qualité afin de signaler les séquences qui contiennent le problème. Et les résultats sont donnés sous forme d'un score de qualité :

- Si le score est entre 0 à 29 le sens est de bonne qualité et la couleur alors est verte.
- Si le score est entre 30 à 99 le sens est de qualité médiocre et la couleur est jaune.
- Si le score est de 100 et plus le sens est de mauvaise qualité et la couleur est rouge.

Sur Nextclade pour le SARS-CoV-2 la paramétrisation est spécifique et les données du CQ sont présentés en vert, leurs désignations sont les suivantes :

N: Données manquantes et s'affiche lorsque plus de 3000 sites n'ont pas été identifiés.

M: Sites mixtes s'affiche lors de la détection des lettres ambiguës.

P: Mutations privées se sont les séquences qui diffèrent entre la séquence de requête et la séquence voisine la plus proche.

C: Amas de mutation s'affiche lors de la détection de plusieurs mutations.

S: Arrêter les codons la réplication des virus ne peut pas avoir de codons d'arrêt prématuré dans les gènes essentiels et de tels arrêts prématurés sont donc un indicateur de séquences problématiques.

F: Déplacements de Frame Nexclade ne traduit pas les séquences de codage à décalage de frame car les insertions ou les suppressions de décalage de Frame entraînent généralement une traduction brouillée ou un arrêt prématuré.

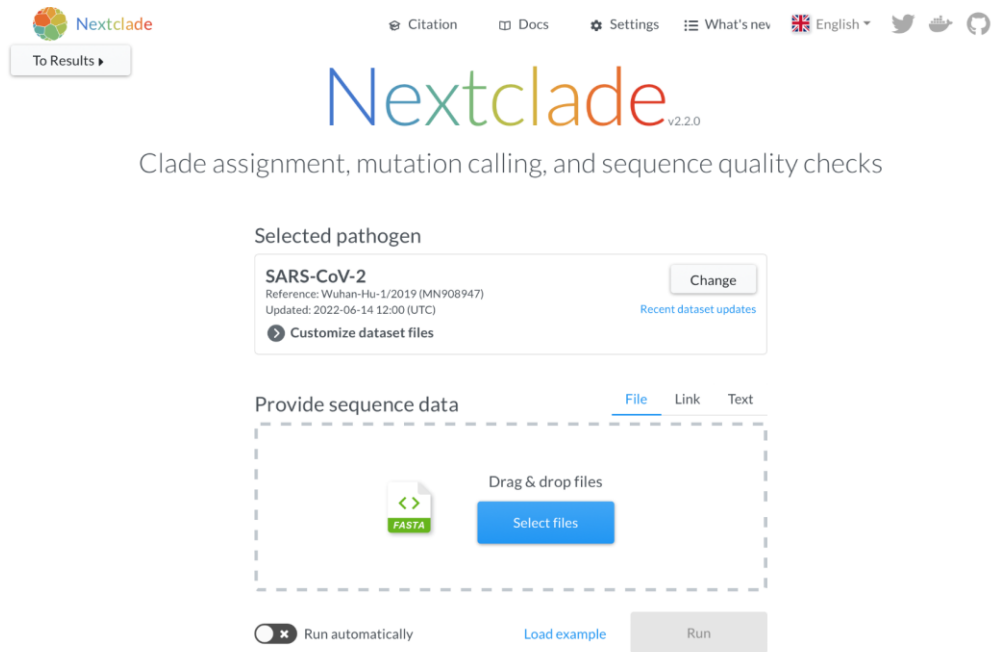


Figure 9: Capture de l'interface de plateforme nexclade.(32)

2.2 Indexation de la référence

On a commencé par trouver le génome de référence du virus SARS-CoV-2 2 avec lequel on doit comparer. Ce génome a été extrait depuis EMBL avec le code MN908947.3. Via le lien suivant : https://www.ebi.ac.uk/ena/browser/view/GCA_009858895.3

L'indexation se fait par le package Minimap2 qui est un programme d'alignement de séquences polyvalent qui aligne les séquences d'ADN ou d'ARNm par rapport à une grande base de données de référence.

```
minimap2 -d (fichier).fasta (fichier).mmi
```

2.3 Alignement de la séquence avec la référence

L'alignement est le processus algorithmique qui consiste à comparer des séquences, nucléotide par nucléotide, pour identifier les similitudes et les différences.

L'outil MINIMAP2 est utilisé pour aligner nos lectures sur le génome de référence et obtenir un fichier SAM (Sequence Alignment/Map), un pour chaque échantillon.

```
minimap2 -ax asm5 Reference.fasta (sequence1).fasta > (aln).sam
```


2.4 Convertir SAM to BAM, trier et indexer le fichier BAM

Pour l'indexation du fichier SAM (SAM to BAM) on a utilisé le logiciel SAMtools, c'est un progiciel pour l'analyse et la manipulation des alignements au format SAM/BAM. Le format BAM est la représentation binaire de SAM, il conserve exactement les mêmes informations que SAM.

Nous convertissons nos fichiers SAM en fichiers BAM, qui sont des versions compressées que de nombreux programmes en aval utilisent comme fichiers d'entrée. On doit d'abord installer **SAMtools** sur la ligne de commande linux Ubuntu, puis exécuter la commande samtools view.

```
samtools view -S -b (aln).sam > (aln).bam
```

Ensuite, on utilise la commande **samtools sort** pour organiser l'alignement mappé par ordre croissant le long des chromosomes du génome de référence.

```
samtools sort (aln).bam -o (aln.sorted).bam
```

Enfin, nous devons créer un index de notre fichier BAM afin que les programmes en aval puissent accéder rapidement à son contenu via la commande suivante :

```
samtools index (aln.sorted).bam
```

2.5 Variant calling

La commande BCFtools est un ensemble d'utilitaires qui manipulent les appels de variantes dans le format d'appel de variantes (VCF) et son homologue binaire BCF. Toutes les commandes fonctionnent de manière transparente avec les VCF et les BCF, on va utiliser deux commandes, la première pour récupérer les fichiers BCF, et la deuxième pour les VCF final.

```
bcftools mpileup -O b -o (sequence1).bcf -f Reference.fasta (aln.sorted).bam
```

```
bcftools call -p ploidy 1 -c -v -o (aln).vcf (sequence1).bcf
```

La première mpileup partie génère des probabilités de génotype à chaque position génomique avec couverture. La deuxième call partie effectue les appels réels

2.6 Annotation

Cette étape annote et prédit les effets des variantes génétiques sur les gènes et les protéines (comme les changements d'acides aminés).

L'outil utilisé est **SnPEff** est la boîte à outils d'annotation de variantes génétiques et de prédiction des effets fonctionnels. <http://snpeff.sourceforge.net/>(33)

```
java -Xmx8g -jar snpEff/snpEff.jar MN908947.3 (aln).vcf > (sequence1).vcf
```

3. Analyse et identification des lignées

Pour l'analyse et l'identification des lignées, nous avons utilisé Pangolin (<https://pangolin.cog-uk.io/>), c'est un outil ou un logiciel qui attribue des lignées aux séquences COVID-19, il stocke des lignées, et organise des données (34).

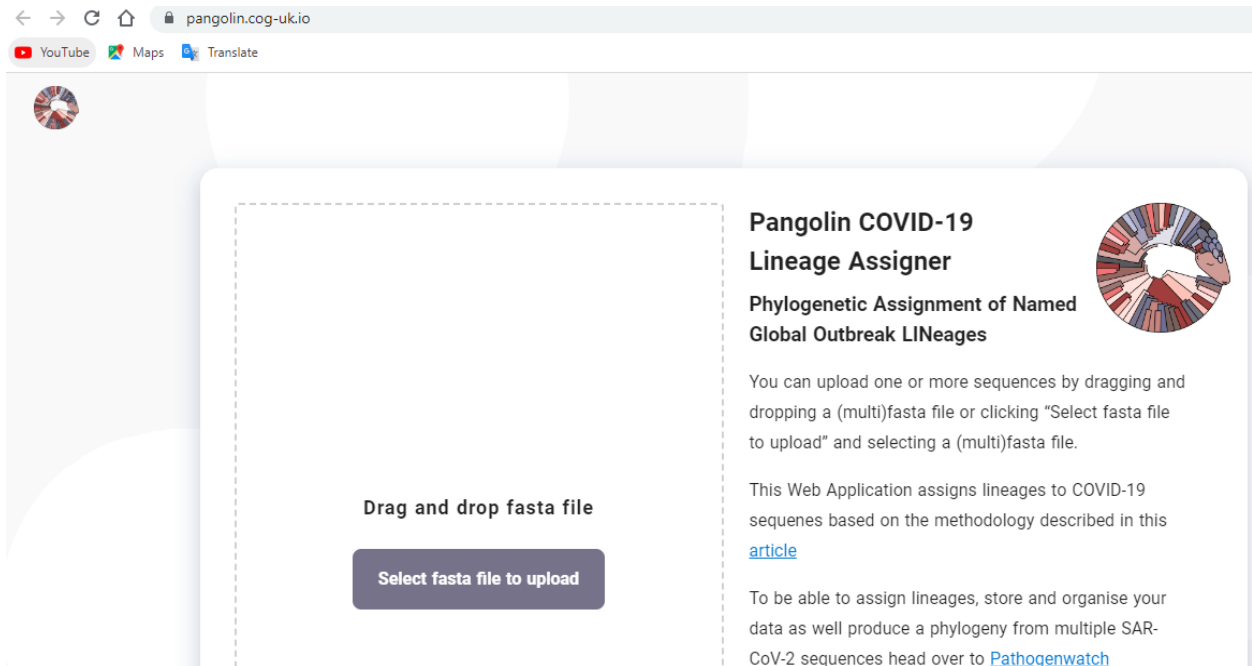


Figure 10: Interface de la plateforme Pangolin.(34)

III. Résultats

1. Contrôle de qualité

Nous avons utilisé un serveur en ligne <https://clades.nextstrain.org/> dans cette étape. On lui a fourni nos séquences sous format FASTA et il nous a retourné un ensemble d'informations sur les séquences y compris « Quality control ». Il vérifie la qualité des séquences en faisant plusieurs contrôles de qualité afin de signaler les séquences qui contiennent le problème.

ID	Sequence name	QC	Clade	Pango lineage (Nextclade)	Mut.	non-ACGTN	Ns	Cov.	Gaps	Ins.	FS	SC
0	hCoV-19/Morocco/FMP-207/2021	N M P C G S	20I (Alpha, V1)	B.1.1.7	35	0	0	99.8%	18	0	0	0(1)
1	hCoV-19/Morocco/FMP-322/2021	N M P C G S	20I (Alpha, V1)	B.1.1.7	40	0	0	99.8%	18	0	0	0(2)
2	hCoV-19/Morocco/FMP-336/2021	N M P C G S	20I (Alpha, V1)	B.1.1.7	36	0	0	99.7%	18	0	0	0(1)

Figure 11: contrôle qualité de la séquence Alpha.

On a divisé nos séquences selon la date de séquençage en quatre groupes, on a fourni les trois séquences du premier groupe au Nextclade, il nous a retourné trois séquences, deux entre eux sont valides, ils ont passé le filtre, et la troisième n'est pas valide, elle n'a pas passé le filtre, car elle est de qualité médiocre.

Pour le reste des séquences, elles ont toutes passés le filtre, elles sont toutes valides. Donc on a choisi de faire l'analyse d'une séquence par groupe.

2. Identification des lignées

Nous avons utilisé Pangolin (<https://pangolin.cog-uk.io/>), pour l'analyse et l'identification des lignées. Nous avons déposé le fichier qui contient nos quatre séquences dans le navigateur web. Pangolin a fait la prédiction des lignées et il nous a spécifié des soulignés des variants auquel appartiennent nos séquences. La page de résultat (Figure12) montre le fichier traité, le nom de la séquence pour chaque séquence et la lignée attribuée.

File name	Sequence name	Lineage	Assignment Conflict
— ANALYSED (Click tick icon for more info) 4 sequences			
✓ sequences.fasta	Alpha1	B.1.1.7	0.0
✓ sequences.fasta	delta1	B.1.617.2	0.0
✓ sequences.fasta	mu1	B.1.621	0.0
✓ sequences.fasta	omicron1	BA.1	0.0

Figure 12: La page de résultats d'identification des lignées sur le site web Pangolin.

Après avoir eu les résultats d'identification des lignées du site Pangolin, on les a téléchargés sous format Excel. D'après le tableau 1, une seule séquence d'omicron appartient à une sous lignée du variant (BA.1), tandis que le reste des séquences correspondent aux lignées principales des variants ; Alpha avec B.1.1.7, Delta avec B1.617.2, Mu avec B.1.621.

Tableau 1: résultat d'identification des lignées des quatre séquences.

Sequence name	Lineage	Scorpio call	Scorpio support
<i>Alpha1</i>	B.1.1.7	Alpha (B.1.1.7-like)	0.96
<i>delta1</i>	B.1.617.2	Delta (B.1.617.2-like)	1
<i>mu1</i>	B.1.621	Mu (B.1.621-like)	0.93
<i>omicron1</i>	BA.1	Omicron (BA.1-like)	0.85

3. Mutation des séquences :

3.1 Alpha :

D'après l'analyse génomique de notre séquence Alpha, on a trouvé qu'elle comporte 38 mutations. La distribution de ces 38 mutations selon le gène est représentée dans la figure 14. On constate que la variante Alpha affecte 44,7% le gène *ARF1ab*, 26,3% le gène *S*, 18,4% le gène *N*, 7,9% le gène *ORF8*, et 2,6% le gène *ORF10*.

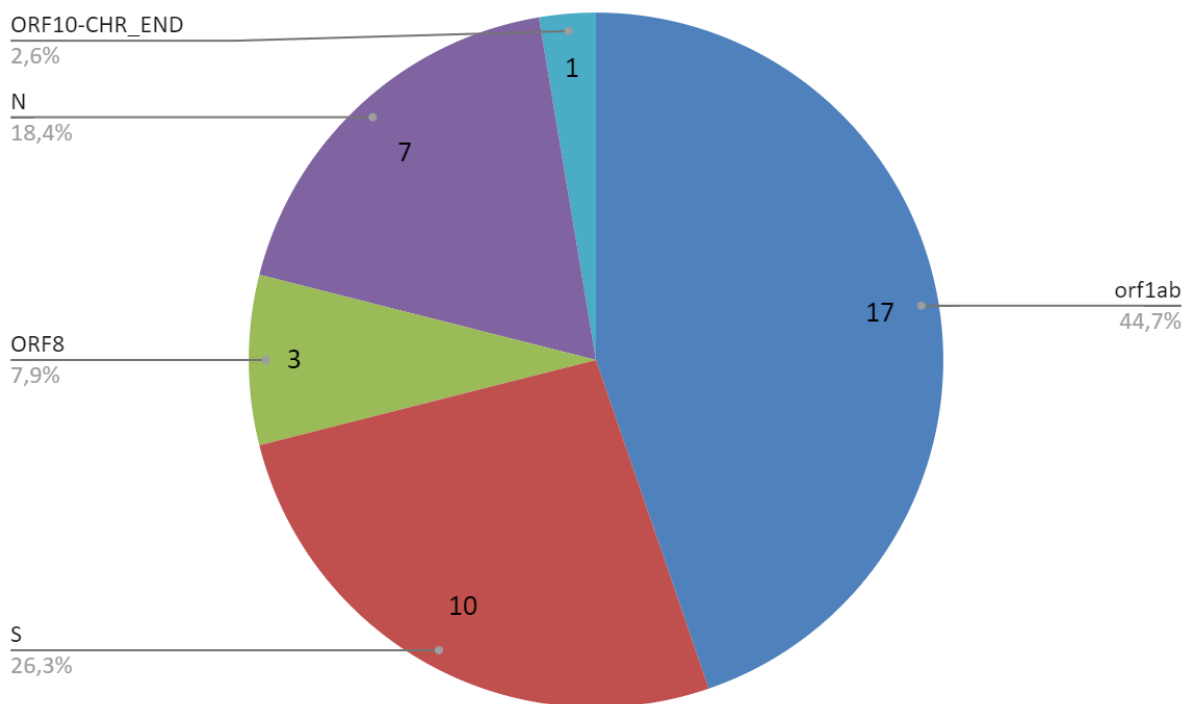


Figure 13: Distribution des mutations de la séquence Alpha.

3 mutations des 38 trouvées sont des délétions, représenté dans le tableau 2. Elles sont toutes codantes et affectent 2 gènes *ORF1ab* et *S*.

Tableau II: : les délétions extraites de la séquence Alpha.

<i>Pos</i>	<i>Réf</i>	<i>Alt</i>	<i>Type</i>	<i>Impact</i>	<i>Gène</i>	<i>Mutation-nt</i>	<i>Mutation-aa</i>
11287	GTCTGG TTTT	G	Conservative inframe	moderate	ORF1ab	11023_11031delTCTGG TTTT	Ser3675_Phe3677del
21764	ATACAT GT	AT	Disruptive inframe	moderate	S	204_209delACATGT	His69_Val70del
21990	TTTATT A	TTT A	Disruptive inframe	moderate	S	432_434delTTA	Y145None

Aucune insertion n'est identifiée dans ces mutations. Parmi ces 38 mutations, nous avons trouvé 35 SNP, dont 2 sont non codantes et 33 codantes. Pour ces derniers, 9 était synonymes et 24 non synonymes.

Concernant les mutations synonymes, elles ont toutes un faible impact et affectent 2 gènes *ORF1ab* et *N*. Pour les mutations non synonymes, elles sont toutes Missense avec un impact non déterminé à l'exception d'une mutation Stop gained qui a un impact élevé. Ces mutations surviennent dans 3 différents gènes avec 10 dans l'*ORF1ab*, 8 dans *S* et 6 dans *N* (tableau 3).

Tableau III: les SNP non synonymes codantes extraites de la séquence Alpha.

Pos	ref	Alt	Type	Impact	gène	Mutation- nt	Mutation- aa
3267	C	T	missense_variant	MODERATE	orf1ab	3002C>T	T1001I
5388	C	A	missense_variant	MODERATE	orf1ab	5123C>A	A1708D
6936	C	T	missense_variant	MODERATE	orf1ab	6671C>T	S2224F
6954	T	C	missense_variant	MODERATE	orf1ab	6689T>C	I2230T
7863	A	G	missense_variant	MODERATE	orf1ab	7598A>G	K2533R
14408	C	T	missense_variant	MODERATE	orf1ab	14144C>T	P4715L
20375	T	C	missense_variant	MODERATE	orf1ab	20111T>C	I6704T
23063	A	T	missense_variant	MODERATE	S	1501A>T	N501Y
23271	C	A	missense_variant	MODERATE	S	1709C>A	A570D
23403	A	G	missense_variant	MODERATE	S	1841A>G	D614G
23604	C	A	missense_variant	MODERATE	S	2042C>A	P681H
23709	C	T	missense_variant	MODERATE	S	2147C>T	T716I
23733	C	T	missense_variant	MODERATE	S	2171C>T	T724I
24506	T	G	missense_variant	MODERATE	S	2944T>G	S982A
24914	G	C	missense_variant	MODERATE	S	3352G>C	D1118H
27972	C	T	stop_gained	HIGH	ORF8	79C>T	Gln27*
28048	G	T	missense_variant	MODERATE	ORF8	155G>T	R52I
28111	A	G	missense_variant	MODERATE	ORF8	218A>G	Y73C
28280	G	C	missense_variant	MODERATE	N	7G>C	D3H
28281	A	T	missense_variant	MODERATE	N	8A>T	D3V
28282	T	A	missense_variant	MODERATE	N	9T>A	D3E
28881	G	A	missense_variant	MODERATE	N	608G>A	R203K
28883	G	C	missense_variant	MODERATE	N	610G>C	G204R
28977	C	T	missense_variant	MODERATE	N	704C>T	S235F

3.2 Delta :

D'après l'analyse génomique de notre séquence Delta, on a trouvé qu'elle comporte 45 mutations. La distribution de ces 45 mutations selon le gène est représentée dans la figure 15, On constate que la variante Delta affecte 40,6% le gène *ORF1ab*, 25,0% le gène *S*, 17% le gène *N*, 9% le gène *ORF7a*, et 6% le gène *ORF3a*. et 3,1% pour les deux gènes *ORF8* et *M*.(35)

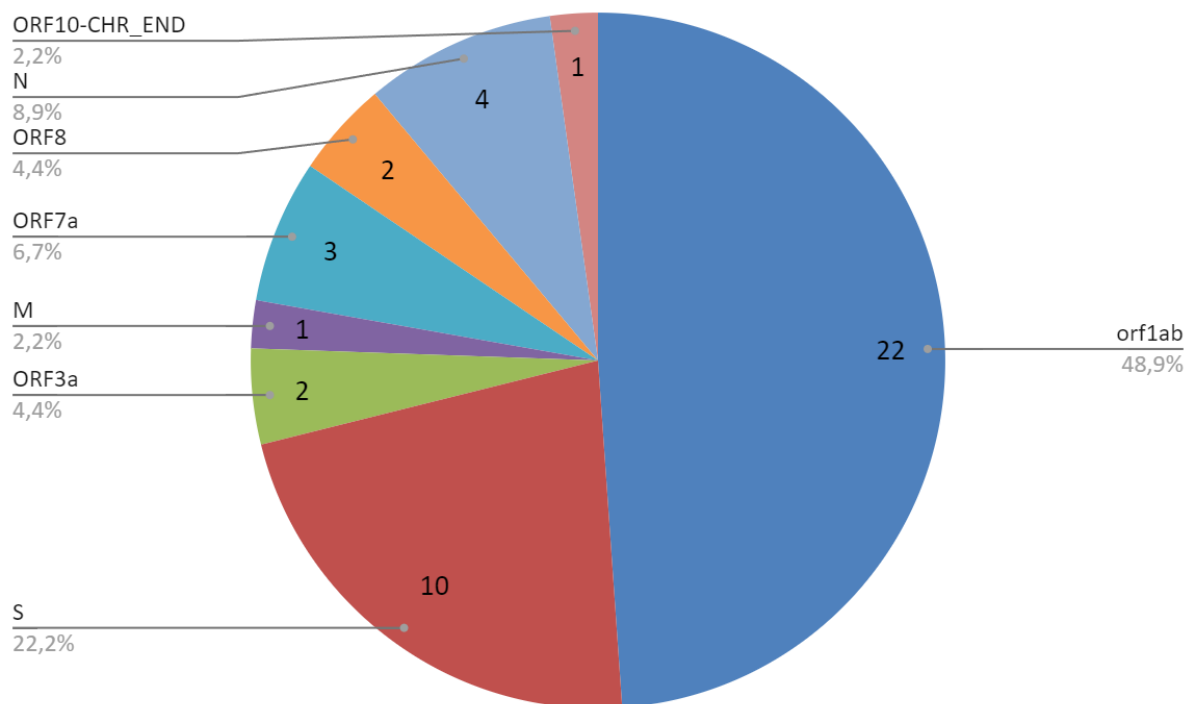


Figure 14: Distribution des mutations par gènes pour la séquence DELTA.

À partir du nombre total de mutations, 2 mutations sont des délétions, représenté dans le tableau 4. Elles sont toutes codantes et affectent un gène *ORF8* et un gène *S*.

Tableau IV: les délétions extraites de la séquence Delta.

Pos	Réf	Alt	type_mutations	Impact	Gène	Mutation-nt	Mutation-aa
22028	GAG TTC A	G	disruptive_inframe_deletion	MODERATE	S	467_472delA GTTCA	Glu156_Arg158delinsGly
28247	AGA TTT C	A	conservative_inframe_deletion	MODERATE	ORF8	355_360delG ATTTC	Asp119_Phe120del

On note qu'aucune insertion n'est présente dans ces mutations. Parmi ces 45 mutations, on a trouvé 43 SNP(36), dont 3 sont non codantes et 40 codantes avec 8 synonymes et 32 non synonymes.

Pour les synonymes, elles ont toute un impact faible et affectent le gène *ORRF1ab* (7 mutations) et le gène *S* (1 mutation).

Pour les non synonymes, elles sont toutes Missense avec un impact non déterminé à l'exception d'une mutation Stop gained qui a un high impact. Ces mutations sont distribuées en 13 affectant le gène *ORRF1ab*, 8 le gène *S*, 2 le gène *ORF3a*, 1 le gène *M*, 3 le gène *ORF7a*, 1 le gène *ORF8* et 4 le gène *N* (tableau 5).

Tableau V: les SNP non synonymes codantes extraites de la séquence Delta.

Pos	Réf	Alt	type_mutations	Impact	Gène	Mutation-nt	Mutation-aa
1191	C	T	missense_variant	MODERATE	orf1ab	926C>T	P309L
5051	C	T	missense_variant	MODERATE	orf1ab	4786C>T	P1596S
5184	C	T	missense_variant	MODERATE	orf1ab	4919C>T	P1640L
6539	C	T	missense_variant	MODERATE	orf1ab	6274C>T	H2092Y
9891	C	T	missense_variant	MODERATE	orf1ab	9626C>T	A3209V
11418	T	C	missense_variant	MODERATE	orf1ab	11153T>C	V3718A
12081	C	T	missense_variant	MODERATE	orf1ab	11816C>T	A3939V
12455	C	A	missense_variant	MODERATE	orf1ab	12190C>A	L4064I
14122	G	A	missense_variant	MODERATE	orf1ab	13858G>A	G4620S
14408	C	T	missense_variant	MODERATE	orf1ab	14144C>T	P4715L
15451	G	A	missense_variant	MODERATE	orf1ab	15187G>A	G5063S
16466	C	T	missense_variant	MODERATE	orf1ab	16202C>T	P5401L
20320	C	T	missense_variant	MODERATE	orf1ab	20056C>T	H6686Y
21618	C	G	missense_variant	MODERATE	S	56C>G	T19R
21987	G	A	missense_variant	MODERATE	S	425G>A	G142D
22917	T	G	missense_variant	MODERATE	S	1355T>G	L452R
22995	C	A	missense_variant	MODERATE	S	1433C>A	T478K
23403	A	G	missense_variant	MODERATE	S	1841A>G	D614G
23604	C	G	missense_variant	MODERATE	S	2042C>G	P681R
24410	G	A	missense_variant	MODERATE	S	2848G>A	D950N
25273	G	T	missense_variant	MODERATE	S	3711G>T	M1237I
25469	C	T	missense_variant	MODERATE	ORF3a	77C>T	S26L
26171	T	A	missense_variant	MODERATE	ORF3a	779T>A	M260K
26767	T	C	missense_variant	MODERATE	M	245T>C	I82T
27638	T	C	missense_variant	MODERATE	ORF7a	245T>C	V82A
27739	C	T	missense_variant	MODERATE	ORF7a	346C>T	L116F
27752	C	T	missense_variant	MODERATE	ORF7a	359C>T	T120I
27972	C	T	stop_gained	HIGH	ORF8	79C>T	Gln27*
28461	A	G	missense_variant	MODERATE	N	188A>G	D63G
28881	G	T	missense_variant	MODERATE	N	608G>T	R203M
29402	G	T	missense_variant	MODERATE	N	1129G>T	D377Y
29427	G	A	missense_variant	MODERATE	N	1154G>A	R385K

3.3 Mu :

D'après l'analyse génomique de notre séquence Mu on a trouvé qu'elle comporte 32 mutations. La distribution de ces 32 mutations selon le gène est représentée dans la figure 16, On constate que la variante Mu affecte 39,1% le gène *ORF1ab*, 43,5% le gène *S*, 8,7% le gène *ORF8*, et 4,3% pour les deux gènes *ORF3a* et *N*.

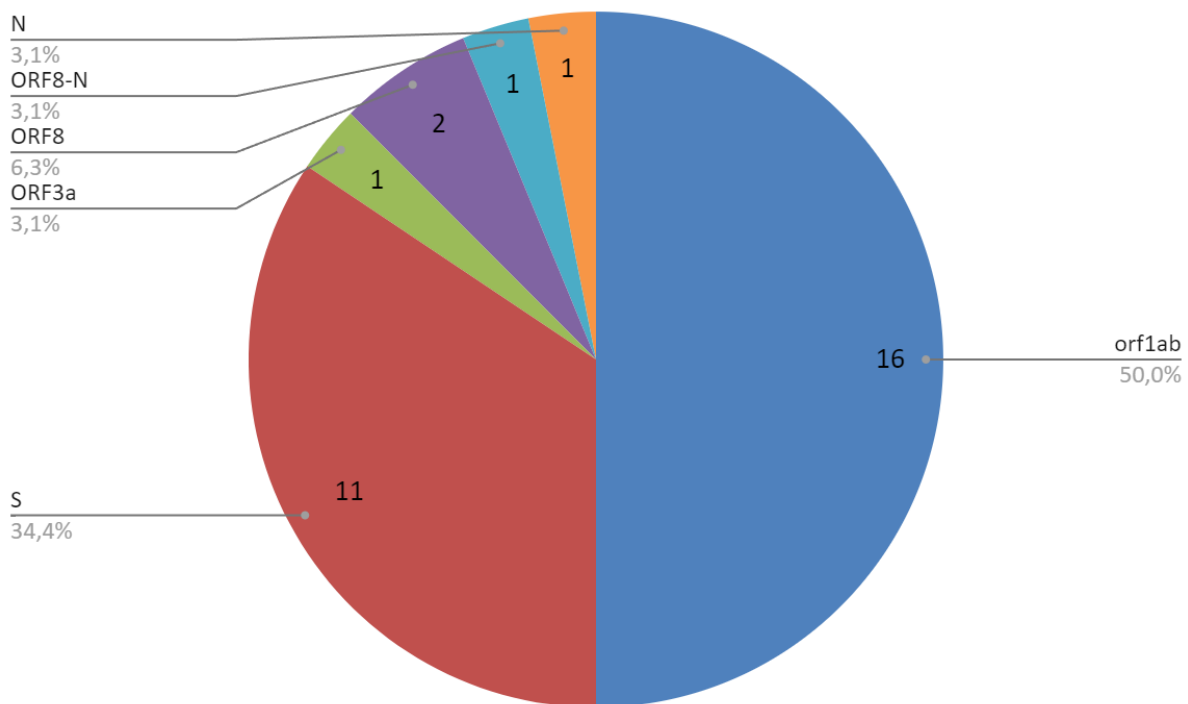


Figure 15: représentation des mutations par gène pour la séquence Mu.

À partir de ces derniers, 1 mutations est une insertion, représenté dans le tableau 6. Cette dernière est codante et touche le gène *S*.

Tableau VI: Les insertions extraites de la séquence Mu.

<i>pos</i>	<i>ref</i>	<i>alt</i>	<i>type_mutations</i>	<i>impact</i>	<i>gene</i>	<i>mutation-nt</i>	<i>mutation-aa</i>
21990	TT	TTACTTCT	Conservative inframe	moderate	S	429_430insACTTCT	Val143_Tyr144insThrSer

Aucune délétion n'est présente dans ces mutations.

Parmi ces 32 mutations, on a trouvé 31 SNP, dont 3 sont non codantes et 28 codantes parmi ces derniers 5 synonymes et 23 non synonymes.

Pour les mutations synonymes, elles sont toute en Low impact et affectent le gène *ORF1ab*.

Quant aux mutations non synonymes, elles sont toutes Missense avec un impact non déterminé, et codent pour 9 protéines dans le gène *ORRF1ab*, 10 dans le gène *S*, et 1 dans chacun des gènes *ORF3a*, *ORF8* et *N* (tableau 7).

Tableau VII: Les SNP non synonymes codantes extraites de la séquence Mu

<i>pos</i>	<i>ref</i>	<i>alt</i>	<i>type_mutations</i>	<i>anno_impact</i>	<i>gene_name</i>	<i>mutation-nt</i>	<i>mutation-aa</i>
1514	C	T	missense_variant	MODERATE	orf1ab	1249C>T	H417Y
3428	A	G	missense_variant	MODERATE	orf1ab	3163A>G	T1055A
4878	C	T	missense_variant	MODERATE	orf1ab	4613C>T	T1538I
10029	C	T	missense_variant	MODERATE	orf1ab	9764C>T	T3255I
11451	A	G	missense_variant	MODERATE	orf1ab	11186A>G	Q3729R
14408	C	T	missense_variant	MODERATE	orf1ab	14144C>T	P4715L
15002	A	G	missense_variant	MODERATE	orf1ab	14738A>G	Y4913C
17019	G	T	missense_variant	MODERATE	orf1ab	16755G>T	E5585D
17491	C	T	missense_variant	MODERATE	orf1ab	17227C>T	P5743S
21846	C	T	missense_variant	MODERATE	S	284C>T	T95I
21993	A	C	missense_variant	MODERATE	S	431A>C	Y144S
21995	T	A	missense_variant	MODERATE	S	433T>A	Y145N
22346	G	T	missense_variant	MODERATE	S	784G>T	A262S
22599	G	A	missense_variant	MODERATE	S	1037G>A	R346K
23012	G	A	missense_variant	MODERATE	S	1450G>A	E484K
23063	A	T	missense_variant	MODERATE	S	1501A>T	N501Y
23403	A	G	missense_variant	MODERATE	S	1841A>G	D614G
23604	C	A	missense_variant	MODERATE	S	2042C>A	P681H
24410	G	A	missense_variant	MODERATE	S	2848G>A	D950N
25563	G	T	missense_variant	MODERATE	ORF3a	171G>T	Q57H
27925	C	A	missense_variant	MODERATE	ORF8	32C>A	T11K
28005	C	T	missense_variant	MODERATE	ORF8	112C>T	P38S
28887	C	T	missense_variant	MODERATE	N	614C>T	T205I

3.4 Omicron :

D'après l'analyse génomique de notre séquence Omicron(37) on a trouvé qu'elle comporte 55 mutations. La distribution de ces 55 mutations selon le gène est représentée dans la figure 17, On constate que la variante Delta affecte 40,0% le gène *ORF1ab*, 40,0% le gène *S*, et 20% le gène *N*.

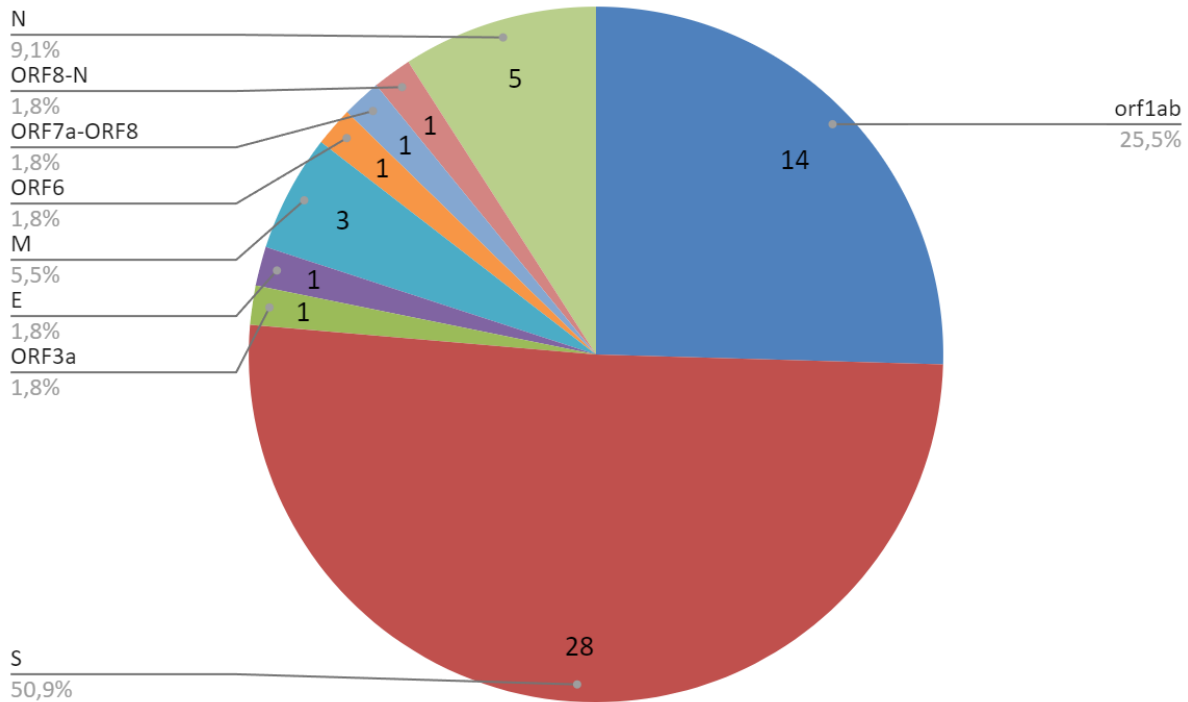


Figure 16: représentation des mutations par gène de la séquence Omicron.

À partir de ces derniers, 5 mutations sont des délétions (représenté dans le tableau 8). Elles sont toutes codantes et affecte les gènes *ORF1ab*, *S* et *N*.(38)

Tableau VIII: les délétions extraites de la séquence Omicron.

<i>pos</i>	<i>ref</i>	<i>alt</i>	<i>type_mutations</i>	<i>anno_i mpact</i>	<i>gene_ name</i>	<i>mutation-nt</i>	<i>mutation-aa</i>
651 2	AGTT	A	disruptive_infra me_deletion	MODER ATE	orf1ab	6248_6250delGT T	Ser2083_Leu20 84delinsIle
112 82	AGTTTGTC TGGTTT	AG TTT	disruptive_infra me_deletion	MODER ATE	orf1ab	11022_11030del GTCTGGTTT	Leu3674_Gly36 76del
217 64	ATACATGT	AT	disruptive_infra me_deletion	MODER ATE	S	204_209delACAT GT	His69_Val70del
219 86	GGTGTTTA TT	G	disruptive_infra me_deletion	MODER ATE	S	425_433delGTGT TTATT	Gly142_Tyr145 delinsAsp
283 61	GGAGAAC GCAG	GG	disruptive_infra me_deletion	MODER ATE	N	90_98delAGAAC GCAG	Glu31_Ser33del

Aucune insertion n'est présente dans ces mutations

Parmi ces 55 mutations on a trouvé 50 SNP, dont 3 sont non codantes et 47 codantes parmi ces derniers 8 synonymes et 39 non synonymes.

Pour les synonymes ils sont toute en Low impact et affectent 4 gènes *ORRF1ab*, un gène *S*, un gène *ORF3a*, un gène *ORF6* et un gène *N*.

Pour les non synonymes ils sont toutes Missense avec un impact non déterminé ces mutations affectent 7 gènes *ORRF1ab*, 25 gènes *S*, un gène *E*, 3 gènes *M*, 3 gènes *ORF7a* et 3 gènes *N* (tableau 9).

Tableau IX: Les SNP non synonymes extraites de la séquence Omicron

pos	ref	alt	type_mutations	anno_impact	gene_name	mutation-nt	mutation-aa
2832	A	G	missense_variant	MODERATE	orf1ab	2567A>G	K856R
8393	G	A	missense_variant	MODERATE	orf1ab	8128G>A	A2710T
10029	C	T	missense_variant	MODERATE	orf1ab	9764C>T	T3255I
10449	C	A	missense_variant	MODERATE	orf1ab	10184C>A	P3395H
11537	A	G	missense_variant	MODERATE	orf1ab	11272A>G	I3758V
14408	C	T	missense_variant	MODERATE	orf1ab	14144C>T	P4715L
18163	A	G	missense_variant	MODERATE	orf1ab	17899A>G	I5967V
21762	C	T	missense_variant	MODERATE	S	200C>T	A67V
22578	G	A	missense_variant	MODERATE	S	1016G>A	G339D
22673	T	C	missense_variant	MODERATE	S	1111T>C	S371P
22674	C	T	missense_variant	MODERATE	S	1112C>T	S371F
22679	T	C	missense_variant	MODERATE	S	1117T>C	S373P
22686	C	T	missense_variant	MODERATE	S	1124C>T	S375F
22813	G	T	missense_variant	MODERATE	S	1251G>T	K417N
22882	T	G	missense_variant	MODERATE	S	1320T>G	N440K
22898	G	A	missense_variant	MODERATE	S	1336G>A	G446S
22992	G	A	missense_variant	MODERATE	S	1430G>A	S477N
22995	C	A	missense_variant	MODERATE	S	1433C>A	T478K
23013	A	C	missense_variant	MODERATE	S	1451A>C	E484A
23040	A	G	missense_variant	MODERATE	S	1478A>G	Q493R
23048	G	A	missense_variant	MODERATE	S	1486G>A	G496S
23202	C	A	missense_variant	MODERATE	S	1640C>A	T547K
23403	A	G	missense_variant	MODERATE	S	1841A>G	D614G
23525	C	T	missense_variant	MODERATE	S	1963C>T	H655Y
23599	T	G	missense_variant	MODERATE	S	2037T>G	N679K
23604	C	A	missense_variant	MODERATE	S	2042C>A	P681H
23854	C	A	missense_variant	MODERATE	S	2292C>A	N764K
23948	G	T	missense_variant	MODERATE	S	2386G>T	D796Y
24130	C	A	missense_variant	MODERATE	S	2568C>A	N856K
24424	A	T	missense_variant	MODERATE	S	2862A>T	Q954H
24469	T	A	missense_variant	MODERATE	S	2907T>A	N969K
24503	C	T	missense_variant	MODERATE	S	2941C>T	L981F
26270	C	T	missense_variant	MODERATE	E	26C>T	T9I
26530	A	G	missense_variant	MODERATE	M	8A>G	D3G
26577	C	G	missense_variant	MODERATE	M	55C>G	Q19E
26709	G	A	missense_variant	MODERATE	M	187G>A	A63T
28311	C	T	missense_variant	MODERATE	N	38C>T	P13L
28881	G	A	missense_variant	MODERATE	N	608G>A	R203K
28883	G	C	missense_variant	MODERATE	N	610G>C	G204R

IV. Discussion

Coronavirus (COVID19) est une maladie infectieuse due au virus SARS-CoV-2 extrêmement contagieuse et très réactive, la première apparition était en chine puis elle s'est propagée dans le monde entier. Ce qui a poussé les pays à analyser les séquences du génome, afin de détecter et confirmer la circulation des variants et de sa propagation, dans cette étude nous avons analysé un ensemble de séquences du SARS-COV 2, en passant par plusieurs étapes, commençant par le contrôle qualité, jusqu'à l'identification des lignées. Nous avons identifié les différentes mutations : (3 délétions et 35 SNP) spécifique à la variante Alpha, (2 délétions et 43 SNP) spécifique à la variante Delta, (1 insertion et 31 SNP) spécifique à la variante Mu, (5 délétions et 50 SNP) spécifique à la variante Omicron. comme comparaison entre le nombre de mutation pour chaque variant, on a trouvé 38 mutations pour Alpha, 45 mutations pour Delta, 32 mutations pour Mu, et 55 mutations pour Omicron (39).

Le variant Omicron alors compte le plus grand nombre de mutations suivi de la variante Delta ensuite la variante Alpha et le dernier et celui de la variante Mu.

Ces mutations sont responsables de la sévérité du virus, sa propagation, et sa gravité. La mutation D614G dans la glycoprotéine S, présente dans les variantes du SARS-CoV-2 Alpha, Delta, Mu et Omicron, (40) a montré l'interaction la plus efficace avec le récepteur ACE2 des cellules. Cette mutation explicite est la principale cause d'infection et de réinfection, de plus elle influence l'immunité et l'évasion partielle du vaccin. Augmente les capacités d'infection, de réplication, et de compétition ainsi que la transmissibilité du virus.

La mutation L452R de Spike, présente dans la variante Delta, permet la liaison du virus aux cellules d'une manière stable, et diminue l'efficacité des vaccins et Anticorps, confère une neutralisation du virus médiée par les anticorps, les mutations L452R peuvent entraîner une forte résistance à l'anticorps monoclonaux.

La mutation N501Y (41) de Spike, présente dans les variantes Alpha et Mu améliore la liaison entre la protéine S et le récepteur ACE2 des cellules cibles. La mutation P681H de Spike présente

dans les variantes Alpha et Mu augmente la production de protéine S par les cellules infectées et favorise la transmission. La mutation E484K de Spike présente dans les variantes Omicron et Mu favorise une résistance à l'immunité. Les mutations qui affectent la protéine Spike la rendent plus active pour pénétrer dans son récepteur et facilitent donc l'entrée du virus dans l'organisme.

V. Conclusion

Le Maroc a connu quatre vagues (Alpha, Delta, Omicron BA1, et Omicron BA5). Dans cette étude nous avons analysé une séquence de chaque variant (Alpha, Delta, Omicron, Mu), en passant par plusieurs étapes, commençant par le contrôle qualité, jusqu'à l'identification des lignées. On a trouvé plusieurs mutations, parmi lesquelles celles qui sont responsables de la sévérité des variants, de la gravité de la maladie, ainsi de leur vitesse de propagation.

La comparaison du nombre de mutation entre variants a révélé que la variante Omicron compte le plus grand nombre de mutation avec 55 mutations suivie de la variante Delta avec 48, Alpha avec 38 et Mu avec 32 mutations. Ce résultat est attendu à cause de l'accumulation des mutations entre les variants et l'ordre d'apparition de ces derniers.

Nous avons aussi identifié le lien entre certaines mutations (l'exemple de D614G, N501Y) et la transmissibilité et propagation du variant. Les mutations donc ont un impact sur les caractéristiques de l'épidémie.

VI. Référence

1. Ketfi A, Chabati O, Chemali S, Mahjoub M, Gharnaout M, Touahri R, et al. Profil clinique, biologique et radiologique des patients Algériens hospitalisés pour COVID-19: données préliminaires. *Pan Afr Med J*. 2020 Jun 15;35(Suppl 2):77.
2. Do the Moroccan SARS-CoV-2 genetic diversity hamper the use of the developed universal vaccines in Morocco? | bioRxiv [Internet]. [cited 2022 Dec 5]. Available from: <https://www.biorxiv.org/content/10.1101/2020.06.30.181123v2>
3. COVID-19 : Modes de transmission et mesures de prévention et de protection contre les risques, incluant le rôle de la ventilation [Internet]. INSPQ. [cited 2022 Dec 5]. Available from: <https://www.inspq.qc.ca/covid-19/environnement/modes-transmission>
4. Helmy YA, Fawzy M, Elasad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J Clin Med*. 2020 Apr 24;9(4):1225.
5. Maladie à coronavirus (COVID-19) [Internet]. [cited 2022 Dec 5]. Available from: <https://www.unicef.org/morocco/recits/maladie-%C3%A0-coronavirus-covid-19>
6. Al-Qahtani AA. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): Emergence, history, basic and clinical aspects. *Saudi J Biol Sci*. 2020 Oct;27(10):2531–8.
7. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. *Genomics*. 2009 Feb;93(2):105–11.
8. Nouveau coronavirus (2019-nCoV): conseils au grand public [Internet]. [cited 2022 Dec 5]. Available from: <https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019/advice-for-public>
9. Beyrampour-Basmenj H, Milani M, Ebrahimi-Kalan A, Ben Taleb Z, Ward KD, Dargahi Abbasabad G, et al. An Overview of the Epidemiologic, Diagnostic and Treatment Approaches of COVID-19: What do We Know? *Public Health Rev*. 2021 Jun 17;42:1604061.
10. CDC. Similarities and Differences between Flu and COVID-19 [Internet]. Centers for Disease Control and Prevention. 2022 [cited 2022 Dec 5]. Available from: <https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm>
11. Abdullahi IN, Emeribe AU, Ajayi OA, Oderinde BS, Amadu DO, Osuji AI. Implications of SARS-CoV-2 genetic diversity and mutations on pathogenicity of the COVID-19 and biomedical interventions. *J Taibah Univ Med Sci*. 2020 Aug;15(4):258–64.

12. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin.* 2020 Sep;41(9):1141–9.
13. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Napoli RD. [Figure, SARS- CoV 2 Structure. Contributed by Rohan Bir Singh, MD; Made with Biorender.com] [Internet]. StatPearls Publishing; 2022 [cited 2022 Dec 5]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK554776/figure/article-52171.image.f3/>
14. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, et al. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol.* 2005 Dec;79(23):14614–21.
15. Enjuanes L, Sola I, Almazan F, Ortego J, Izeta A, Gonzalez JM, et al. Coronavirus derived expression systems. *J Biotechnol.* 2001 Jul 12;88(3):183–204.
16. Risco C, Antón IM, Enjuanes L, Carrascosa JL. The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *J Virol.* 1996 Jul;70(7):4773–7.
17. Histoire du COVID-19 – C5 : Infectiosité et réplication hors norme du SARS-Cov2 [Internet]. FranceSoir. [cited 2022 Dec 5]. Available from: <https://www.francesoir.fr/societe-science-tech/histoire-du-covid-19-c5-infectiosite-et-replication-hors-norme-du-sars-cov2>
18. Jia HP, Look DC, Shi L, Hickey M, Pewe L, Netland J, et al. ACE2 Receptor Expression and Severe Acute Respiratory Syndrome Coronavirus Infection Depend on Differentiation of Human Airway Epithelia. *J Virol.* 2005 Dec 15;79(23):14614–21.
19. Zhou Z, Qiu Y, Ge X. The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Anim Dis.* 2021;1(1):5.
20. Laamarti M, Essabbar A, Alouane T, Kartti S, Boumajdi N, Bendani H, et al. Do the Moroccan SARS-CoV-2 genetic diversity hamper the use of the developed universal vaccines in Morocco? [Internet]. bioRxiv; 2020 [cited 2022 Dec 5]. p. 2020.06.30.181123. Available from: <https://www.biorxiv.org/content/10.1101/2020.06.30.181123v2>
21. Mittal A, Manjunath K, Ranjan RK, Kaushik S, Kumar S, Verma V. COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. *PLOS Pathog.* 2020 Aug 21;16(8):e1008762.
22. Morse JS, Lalonde T, Xu S, Liu WR. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *Chembiochem Eur J Chem Biol.* 2020 Mar 2;21(5):730–8.
23. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. *Eur Arch Oto-Rhino-Laryngol Off J Eur Fed Oto-Rhino-Laryngol Soc EUFOS Affil Ger Soc Oto-Rhino-Laryngol - Head Neck Surg.* 2020 Aug;277(8):2251–61.

24. Nicas M, Best D. A study quantifying the hand-to-face contact rate and its potential application to predicting respiratory tract infection. *J Occup Environ Hyg.* 2008 Jun;5(6):347–52.
25. Tu YF, Chien CS, Yarmishyn AA, Lin YY, Luo YH, Lin YT, et al. A Review of SARS-CoV-2 and the Ongoing Clinical Trials. *Int J Mol Sci.* 2020 Apr 10;21(7):2657.
26. Kannan SR, Spratt AN, Cohen AR, Naqvi SH, Chand HS, Quinn TP, et al. Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *J Autoimmun.* 2021 Nov;124:102715.
27. Daum LT, Rodriguez JD, Worthy SA, Ismail NA, Omar SV, Dreyer AW, et al. Next-generation ion torrent sequencing of drug resistance mutations in *Mycobacterium tuberculosis* strains. *J Clin Microbiol.* 2012 Dec;50(12):3831–7.
28. CEPI’s collaborative task force to assess COVID-19 vaccines on emerging viral strains [Internet]. [cited 2022 Dec 5]. Available from: <https://www.biospectrumasia.com/news/83/17161/cepi-collaborative-task-force-to-assess-covid-19-vaccines-on-emerging-viral-strains.html>
29. Galindo-González L, Pinzón-Latorre D, Bergen EA, Jensen DC, Deyholos MK. Ion Torrent sequencing as a tool for mutation discovery in the flax (*Linum usitatissimum* L.) genome. *Plant Methods.* 2015;11:19.
30. Damerla RR, Chatterjee B, Li Y, Francis RJB, Fatakia SN, Lo CW. Ion Torrent sequencing for conducting genome wide scans for mutation mapping analysis. *Mamm Genome Off J Int Mamm Genome Soc.* 2014 Apr;25(0):120–8.
31. Wei J, Li Y. Airborne spread of infectious agents in the indoor environment. *Am J Infect Control.* 2016 Sep 2;44(9 Suppl):S102-108.
32. Nextclade [Internet]. [cited 2022 Dec 5]. Available from: <https://clades.nextstrain.org>
33. SnpEff and SnpSift [Internet]. [cited 2022 Dec 5]. Available from: <http://pcingola.github.io/SnpEff/>
34. COG-UK [Internet]. [cited 2022 Dec 5]. Available from: <https://pangolin.cog-uk.io/>
35. variant delta Archives [Internet]. Telquel.ma. [cited 2022 Dec 5]. Available from: <https://telquel.ma>
36. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinforma Oxf Engl.* 2011 Nov 1;27(21):2987–93.
37. Le variant “Omicron” a des particularités rares, dont de nombreuses mutations (Professeur) | MapNews [Internet]. [cited 2022 Dec 5]. Available from: <http://www.mapnews.ma/fr/actualites/social/le-variant-omicron-des-particularit%C3%A9s-rares-dont-de-nombreuses-mutations>

38. Araf Y, Akter F, Tang YD, Fatemi R, Parvez MSA, Zheng C, et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *J Med Virol*. 2022 May;94(5):1825–32.
39. Alouane T, Laamarti M, Essabbar A, Hakmi M, Bouricha EM, Chemaoui-Elfihri MW, et al. Genomic Diversity and Hotspot Mutations in 30,983 SARS-CoV-2 Genomes: Moving Toward a Universal Vaccine for the “Confined Virus”? *Pathogens*. 2020 Oct 10;9(10):829.
40. Laghdaf SM, El Baraa A, Mohamed Beydjeu A. Caractérisation d’un cluster des cas de COVID-19 liés au variant Omicron, en Mauritanie. *Tunis Médicale*. 2022 Mar;100(3):217–21.
41. COVID19 : mutations, variants, lignées, N501Y, E484K... de quoi parlez-vous ? [Internet]. VIDAL. [cited 2022 Dec 5]. Available from: <https://www.vidal.fr/actualites/26672-covid-19-mutations-variants-lignes-n501y-e484k-de-quoi-parlez-vous.html>
42. Ren SY, Wang WB, Gao RD, Zhou AM. Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. *World J Clin Cases*. 2022 Jan 7;10(1):1–11.
43. Syed AM, Ciling A, Khalid MM, Sreekumar B, Chen PY, Kumar GR, et al. Omicron mutations enhance infectivity and reduce antibody neutralization of SARS-CoV-2 virus-like particles. *MedRxiv Prepr Serv Health Sci*. 2022 Jan 2;2021.12.20.21268048.