Royaume du Maroc
Ministère de l'Éducation Nationale, de la Formation Professionnelle, de L'Enseignement Supérieur et de la Recherche
Université Mohammed V de RABAT
Faculté de médecine et de pharmacie de rabat

Année : 2022 ***N° : MM122022***

MEMOIRE DE MASTER

MASTER DE BIOTECHNOLOGIE MEDICALE

OPTION : Biomédicale

Thème

Etude génomique de la variante omicron du SARS-COV-2 au Maroc

Présentée par :
Belhaj Sara

Encadrée par :
Pr.Ouadghiri Mouna

Jury de soutenance : **Date de soutenance :20/10/2022**

Pr. OUADGHIRI Mouna Faculté de médecine et de pharmacie **: Encadrante**

Pr. LOUATI Sara, Faculté de médecine et de pharmacie **: Présidente**

Pr.KANDOUSSI Ilham, Faculté de médecine et de pharmacie **: Examinatrice**

Remerciements

Aux membres du jury :

Je remercie sincèrement madame la professeure LOUATI. Sara de me faire l'honneur de présider la soutenance de mon PFE.

Je remercie également les membres du jury, mesdames les Professeurs : OUADGHIRI.Mouna, KANDOUSSI. Ilham et Dr. BENDANI. Houda. Je vous remercie de m'avoir fait l'honneur d'accepter de faire partie de mon jury de mémoire, qu'il me soit permis, de vous exprimer toute ma reconnaissance, mon respect et mon estime.

Veuillez croire à l'expression de mes sentiments les plus distingués.

À mes directeurs de mémoire :

J'exprime ma plus vive gratitude à ma directrice de mémoire, madame la Professeur OUADGHIRI. Mouna et ma co-directrice, madame le Docteur Bendani Houda, de m'avoir accompagnée et soutenue tout au long de ce travail. Votre expertise m'a été très précieuse.

À toute l'équipe du Medbiotech Que reçoivent également l'expression de ma profonde gratitude : Tous les professeurs du Master pour les cours enseignés tout au long de cette formation.

Je remercie également Mes collègues du groupe, pour tous les moments confraternels passés ensemble durant notre formation. Merci pour votre belle humeur et vos sourires, j'espère que notre groupe va perdurer en dehors de la faculté.

Dédicaces

À ma famille :

Merci à mes parents, mes sœurs et mon frère de me soutenir dans chacun de mes pas et à chaque étape de ma vie. J'espère vous rendre fiers. J'exprime également ma reconnaissance à ma belle-famille pour son soutien et son accueil chaleureux. Merci à mon mari d'être là, chaque jour auprès de moi, et de m'avoir soutenue tout au long de cette formation, et à ma fille qui illumine ma vie chaque jour.

À mémoire de ma belle-mère Qu'elle repose en paix.

Résumé

En 2020 le monde a été bouleversé par la survenue d'une pandémie liée à un nouveau coronavirus : le SARS-CoV-2, de la maladie à Coronavirus 2019 (COVID-19), qui a couté la vie à plus de 1,8 millions de personnes dans le monde à la fin de l'année. L'Organisation Mondiale de la Santé a annoncé le 30 janvier 2020 que cette nouvelle épidémie de maladie COVID-19 était classée comme une « urgence de santé publique » de portée internationale. La communauté médicale s'est depuis mobilisée pour lutter contre cette infection virale et a dû, à marche forcée, se familiariser avec cette nouvelle maladie. Certains aspects virologiques, cliniques et thérapeutiques sont rappelés ici.

On a établi lors de ce travail une analyse génomique d'une séquence d'omicron au Maroc ensuite on a réalisé une analyse phylogénétique avec 864 séquences (52 marocaine) afin de comprendre l'évolution du variant au Maroc et son origine. Le but de cette étude repose étroitement sur l'analyse de la séquence marocaine BA.1.

Lors de notre analyse de la séquence marocaine BA.1, on a identifié un total de 60 mutations avec 6 délétions et 43 non synonymes SNP (polymorphisme d'un seul nucléotide). Ce variant se distingue des autres variantes avec 32 mutations dans la protéine spike dont l'impact varie entre l'augmentation de la pathogénicité, de la transmission et/ou de la virulence du virus. L'analyse phylogénétique a regroupé les séquences du Maroc en deux différents clades ; un clade de séquence marocaine qui exprime l'évolution interne du variant au Maroc et des clades hétérogènes montrant que le profil de la variante BA.1 au Maroc est importé.

Mots clés : variante BA.1, Omicron, SNP, phylogénétique, Sars-CoV-2

Abstract

In 2020 the world was turned upside down by the occurrence of a pandemic linked to a new coronavirus: SARS-CoV-2, from Coronavirus disease 2019 (COVID-19), which cost the lives of more than 1.8 million people worldwide by the end of the year. The Organization World Health announced on January 30, 2020 that this new epidemic of disease COVID-19 was classified as a "public health emergency" of international concern. The medical community has since mobilized to fight this viral infection and has forced to familiarize themselves with this new disease. Some aspects virological, clinical and therapeutic are recalled here. During this work, a genomic analysis of an omicron sequence in Morocco was established. Then we carried out a phylogenetic analysis with 864 sequences (52 Moroccan) in order to understand the evolution of the variant in Morocco and its origin. The aim of this study is based closely on the analysis of the Moroccan sequence BA.1. During our analysis of the Moroccan BA.1 sequence, we identified a total of 60 mutations with 6 deletions and 43 non-synonymous SNPs (single nucleotide polymorphism). This variant stands out from other variants with 32 mutations in the spike protein whose impact varies between increased pathogenicity, transmission and/or virulence of the virus. The phylogenetic analysis grouped the sequences from Morocco into two different clades; a Moroccan sequence clade that expresses the internal evolution of the variant in Morocco and the heterogeneous clades showing that the profile of the BA.1 variant in Morocco is imported

Keywords: BA.1 variant, Omicron, SNP, Phylogenetics, Sars-CoV-2.

ملخص

في عام 2020 عرف العالم جائحة مرتبطة بفيروس كورونا الجديد: سارس كوف- 2، المسؤول عن مرض فيروس كورونا 2019 (كوفيد -19)، الذي أودى بحياة أكثر من 1.8 مليون شخص في جميع أنحاء العالم. في نهاية العام أعلنت منظمة الصحة العالمية عن تفشي مرض كوفيد -19 حيث تم تصنيفه على أنه "حالة طوارئ صحية عامة" مما أثار قلقا دوليا. منذ ذلك الحين، تجند المجتمع الطبي لمكافحة هذه العدوى الفيروسية للتعرف على هذا المرض الجديد. من خلال هذا العمل، قمنا بتحليل لتطور الوراثي لأزيد من 864 عينة (مع 52 من أصل مغربي). وذلك من أجل فهم تطور هذا المتغير والكشف عن أصله، الهدف من هذه الدراسة يرتكز أساسا على تحليل العينة ب أ1 ذات الأصل المغربي.

بعد ذلك تمكنا من تحديد 60 طفرة مع 6 عمليات حذف و43، كما كشفنا عن اختلاف المتغير عن باقي المتغيرات الأخرى والتي تحتوي على 32 طفرة في بروتين سبايك التي يختلف تأثيرها بين زيادة الإمراضية و/أو انتقال و / أو ضراوة الفيروس. من خلال جمع التحليل الجيني العينات ذات الأصل المغربي في مجموعتين مختلفتين. بحيث أن المجموعة الأولى تضم فقط عينات من أصل مغربي وذك للكشف عن التطور الداخلي للمتغير داخل المغرب أما بالنسبة للمجموعة الثانية فهي تضم عينات من مختلف البلدان حيث تبين أن المتغير ب أ1مستورد.

الكلمات الدالة: متحور، فيروس كورونا 2019, طفرة, سارس كوف- 2 ، تحليل جيني.

Liste des figures

LISTE DES ANNEXES FIGURES

Liste des tableaux

Liste des abréviations

ACE2 : Enzyme de conversion de l'angiotensine 2

ADN : Acide désoxyribonucléique

ARN : Acide ribonucléique

BAM: BAM File Format

BWA: Burrows-Wheeler Alignment

CE : Electrophorèse capillaire

COVID-19 : Maladie du coronavirus 2019

CoV: Coronavirus

CMOS: Complementary metal oxide semi-conductor

CT-Scan: Computed Tomography (CT) Scan

DNTp: désoxyribonucléotide triphosphate

DPI: Dots Per Inch

E: Protéine d'enveloppe

FIV: fécondation in vitro

GISAID: Global Initiative on Sharing Avian Influenza Data

IA : intelligence artificielle

IgG/IgM : immunoglobulines de type G/M

IL : Interleukine

IRM : imagerie par résonance magnétique

ISFET: Ion Sensitive Field Effect Transistor

KDa : kilo dalton

M : Protéine de membrane

MAFFT: Multiple Alignment using Fast Fourier Transform

MERS-CoV : Syndrome respiratoire du Moyen-Orient

MSA : Multiple sequence alignment

N : Protéine de nucléocapside

NGS: next-generation sequencing

NLRP3: NOD-like receptor family, pyrin domain containing 3

NSP : Non structural protein (Protéine non structurale)

OMS : Organisation mondiale de la santé

ORF : open reading frame (Cadre de lecture ouvert)

Pb : paire de bases

PCR: Polymerase Chain Reaction

PM: Pico moles

Phyml: Phylogenetic estimation using Maximum Likelihood

RBD : Domaine de liaison au récepteur

RdRP : RNA-dependent RNA polymerase

RT-PCR : Réaction de polymérisation en chaîne par transcription inverse

RT-qPCR : reverse transcription Polymerase Chain Reaction

S : Protéine Spike

SAM : sequence alignment map

SARS-CoV-2 : Coronavirus 2 du syndrome respiratoire aigu sévère

SBS : Sequencing By Synthesis

SDRA : Syndrome de détresse respiratoire aiguë

SMS: smart model selection

SNP: Single Nucleotide Polymorphism

TDM : La tomodensitométrie

TMPRSS2 : Protéase transmembranaire à sérine 2

VCF : Variant Call Format

VIH : virus de l'immunodéficience humaine

WGS : whole genome sequencing

Table des matières

Introduction

Le covid-19 poursuit son expansion dans le monde entier ce qui a abouti à des mutations influençant sur la facilité de sa propagation. La pandémie de Covid-19 qui depuis plus de deux ans affecte la planète a mis en évidence, de grandes disparités entre pays en matière de séquençage qui est une technique indispensable pour surveiller la circulation du SARS-CoV-2 et son évolution. Par conséquent, son évolution doit être surveillée de près. En effet, comme tous les virus à ARN, le SRAS-CoV-2 tend fortement à muter.(1) Les nombreuses variantes successives – Alpha, Beta, Gamma, Delta et finalement Omicron – démontrent la capacité du virus à s'adapter rapidement. Les variantes n'ont pas la même transmissibilité et la même pathogénicité que la souche sauvage du virus. Il est donc indispensable de suivre leur évolution pour anticiper les risques. Il importe tout autant de caractériser ces variantes afin d'évaluer l'efficacité des vaccins et des traitements.(2) Le meilleur outil disponible pour l'identification de nouvelles mutations préoccupantes ainsi que le suivi de l'évolution du SARS-CoV-2 demeure le séquençage.(3)

La surveillance génomique est une discipline associant l'épidémiologie et la phylogénie. Le séquençage à grande échelle des génomes, largement répondu aujourd'hui, peut servir à lier l'évolution des séquences au développement épidémique. Afin de comprendre la biologie du SARS-CoV-2 afin d'apercevoir des stratégies thérapeutiques ou préventives capables de contenir la pandémie. Pour cette raison, il faut impérativement d'identifier les mutations observées dans la séquence génomique du SARS-CoV-2 et de déterminer si elles peuvent être utilisées pour fournir une indication pour l'aptitude et l'adaptation virale. En effet, il est possible que les variantes mutationnelles modulent la propagation de la maladie, également la présentation clinique du COVID-19.(4)

Dans le cadre de notre étude, nous allons nous focaliser sur la variante Omicron, voir sa propagation et son évolution. Les données rapportées ici sont le reflet de l'état des connaissances du début de cette vague jusqu'à la fin juillet 2022.

Nous portons notre attention principalement sur l'analyse des séquences d'Omicron du génome viral pour déterminer les différentes caractéristiques de ce variant. Une analyse phylogénétique et évolutive approfondie nous permettra d'apporter des informations importantes sur l'évolution du SARS-CoV-2 entraînée par la propagation virale et la pression sélective.

L'intérêt primordial de ce travail est l'étude de la variante Omicron ainsi de déterminer et préciser la relation entre la sévérité de ce variant et leurs mutations extraites à travers l'analyse génomique de la séquence étudiée. Nous allons tout d'abord analyser une séquence Omicron du Maroc et extraire les mutations, puis nous allons les comparer aux mutations circulant dans le monde. Une étude phylogénétique de certaines séquences d'omicron issue de divers pays a été utilisé pour voir l'évolution et la transmission de la variante.

I La pandémie de covid-19

1 Le coronavirus : généralités

1.1 Éléments historiques

Le coronavirus a été découvert en 1967 par Tyrrell et a été nommé en raison de projections en forme de couronne de sa surface. En 1930, le coronavirus a regroupé à partir des critères morphologiques des virus infectant les animaux et les hommes. En 2003, une population chinoise dans la province du Guangdong a été infectée par un virus causant le syndrome respiratoire aigu sévère (SARS), ce virus a été nommé le SARS-COV (le coronavirus du syndrome respiratoire aigu sévère). Les patients infectés par ce virus présentaient un syndrome de détresse respiratoire aiguë (SDRA)(5). En 2012, deux ressortissants saoudiens ont été infectés par un autre coronavirus nommé Coronavirus du Syndrome Respiratoire du Moyen-Orient (MERS-COV) qui est un membre du sous-groupe bêta coronavirus. L'infection par le MERS-COV commence à partir d'une légère lésion des voies respiratoires supérieures tandis que la progression conduit à une maladie respiratoire grave. Les patients infectés par le MERS-coronavirus souffrent aussi d'un syndrome de détresse respiratoire aiguë (SDRA)(1).

En décembre 2019, une épidémie de pneumonies, décrite à l'époque comme d'allure virale de cause inconnue a émergé dans la ville de Wuhan (province de Hubei en chine).

Les premiers cas étaient soit des propriétaires des magasins, soit des personnes qui avaient visité le marché des fruits de mer, pendant que certaines personnes avaient capté l'infection même sans avoir visité le marché des fruits de mer. Ces observations indiquent qu'il se transmet par contact rapproché avec des personnes infectées (1).

Le 9 janvier 2020, la découverte d'un nouveau coronavirus a été annoncée officiellement par les autorités sanitaires chinoises et l'OMS. D'abord appelé 2019-nCOV puis SARS-COV-2, ce nouveau virus est l'agent responsable de cette nouvelle maladie infectieuse respiratoire appelée covid-19 (4).

1.2 Taxonomie

Les coronavirus (COV) sont divisés en quatre genres, dont α / β / γ / δ-COV. Le α et le β-COV sont capables d'infecter les mammifères, tandis que le γ- et le δ-COV ont l'aptitude à infecter les oiseaux. Ces COV appartiennent au sous-ordre des cornidovirineae au sein de l'ordre des Nidovirales et de la famille des Coronaviridae. Certains coronavirus sont connus par leurs capacités d'infecter les humains, deux d'entre eux appartiennent aux Alpha

coronavirus : HCOV-229E, HCoV-NL63, et quatre virus appartiennent aux Bêta-coronavirus HCOV-OC43, HCOV-HKU1, et les deux virus mortels SARS-COV et MERS-COV. Alors que Le SARS-COV-2 est classifié par la 7ème souche de coronavirus pathogène pour l'homme et qui est responsable de la maladie du COVID-19 (2).

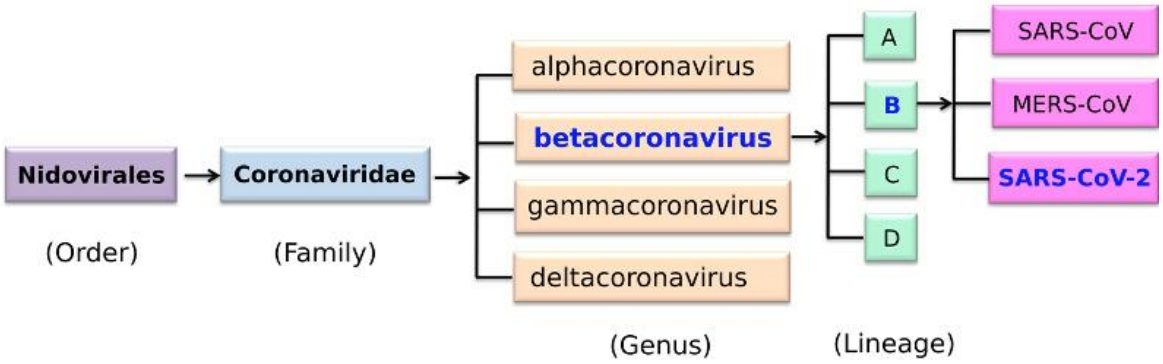


Figure 1: Classification taxonomique du coronavirus (6).

1.3 **Epidémiologie :**

La maladie du COVID-19 est une infection respiratoire aiguë très contagieuse due à un virus zoonotique appelé (SARS-CoV-2), apparue depuis décembre 2019 à Wuhan, en Chine, et qui a rapidement acquis la capacité de transmission interhumaine ce qui aboutit à une épidémie importante en Chine, avec de nombreux cas importés enregistrés dans différents pays à travers le monde. L'Organisation Mondiale de la Santé (OMS) a déclaré, le 30 janvier 2020, que cette épidémie constituait une Urgence de Santé Publique de Portée internationale, conformément aux dispositions du Règlement Sanitaire International, L'analyse des résultats de ces articles menés dans le but d'identifier les facteurs génétiques associés à la gravité de la maladie SARS-COV-2, ainsi la détermination de mutations déterminer liés à la sévérité de la maladie COVID-19.

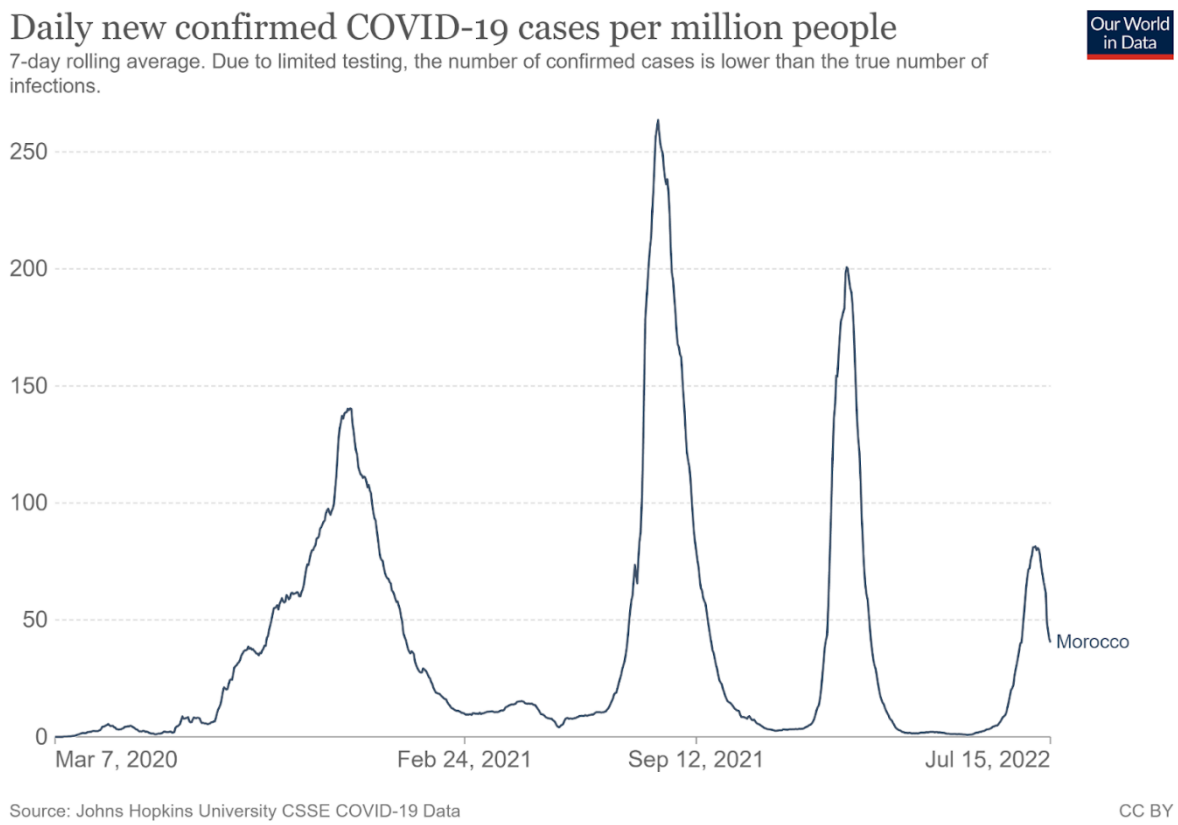


Figure 2: Statistiques relatives à l'évolution de l'épidémie depuis son émergence au Maroc (7).

Depuis l'apparition du nouveau coronavirus en Chine en décembre 2019, la propagation de cette maladie inconnue n'a cessé de s'étendre dans le monde, induisant à une crise sanitaire et mettant en quarantaine plus de la moitié de la population. Plus de 23 millions des cas de COVID-19 ont été signalés et la maladie a fait plus de 802 693 victimes à l'échelle mondiale le 23 août 2020.

Ce graphique nous présente des statistiques indiquant le développement d'épidémie depuis son apparition au Maroc. Le 2 mars 2020, le Maroc a enregistré le premier cas de COVID-19. On observe que le nombre des cas confirmés a augmenté progressivement ce qui a amené notre pays à mettre des mesures de distanciation en œuvre consistant en confinement depuis le 20 mars ce qui a permis un ralentissement de la propagation d'infection avec le variant Alpha. Simultanément à la diffusion du variant Alpha, d'autres variantes significatives ont été détectés lors de la deuxième vague du covid-19 on trouve le variant Delta qui est apparu au Maroc en avril 2021, qui s'est propagé rapidement et est devenu le variant le plus dominant. Ensuite, en octobre, on est revenu à une courbe de contamination qui s'est aplatie et un niveau d'incidence et de positivité très bas, ce qui correspond à une période où le niveau de transmission est très bas et circule à un faible niveau et non pas sous forme de vague, en parlant donc d'une inter-vague.

Les cas de covid-19 reliés au variant Omicron représente la 3éme vague de transmission communautaire du virus, qui a assuré le rapprochement du pic entre le 3 et 17 janvier 2022, alors que le pic de la vague d'Omicron a été enregistré du 17 au 23 janvier 2022. Après 5 semaines de hausse continue, le Maroc a connu une diminution de cas infectés par le virus au cours de la semaine du 24 au 30 janvier, ce qui a indiqué que le Maroc est dans la fin de de la vague omicron et a entamé la phase post vague qui désigne la phase de déclin. Donc la vague Omicron était court, elle a durée 11 semaines et a atteint son apogée dans la semaine du 17 au 23 janvier 2022, Comparé à la vague "Delta", la vague Omicron est moins virulent et moins létal.

La situation épidémiologique relative au Covid-19 au Maroc demeure stable depuis le mois février de cette année, on assiste à l'augmentation du taux de transmission par rapport au sous-variant Omicron BA.2 qui est dominant actuellement au Maroc. La famille Omicron continue de s'agrandir donc on peut conclure que la pandémie de la Covid-19 est loin d'être terminée. Alors que le Maroc fait face à une quatrième vague de contaminations portée par le BA.5.

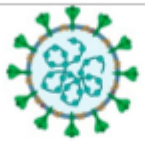
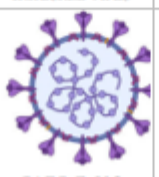
	Pathogène	Numéro de reproduction de base ou R0	Taux de létalité ou TL	Période d'incubation	Taux d'hospitalisation	Taux d'attaque de la communauté	Nombres des infectés annuelles globalement
SARS	 SARS-CoV	3	9,6-11%	2-7 jours	La plupart des cas	10-60%	8098 (à 2003)
MERS	 MERS-CoV	0,3-0,8	34,4 %	6 jours	La plupart des cas	4-13%	420
FLU	 Influenza virus	1,3	0,05-0,1%	1-4 jours	2%	10-20%	1 milliard
Covid-19	 SARS-CoV-2	2,0-2,5	3,4%	4-14 jours	19%	30-40%	indisponible

Tableau 1: Comparaison épidémiologique des infections virales respiratoire (8).

Un aperçu des diverses statistiques épidémiologiques associées aux épidémies de grippe saisonnière, de SARS, de MERS et de COVID-19 est présenté dans le tableau ci-dessus. Une recherche a montré que tous les âges sont sensibles au COVID-19. Notant que les patients symptomatiques et asymptomatiques transmettent la maladie. Des études ont démontré que la charge virale n'est pas significative peu différente entre les individus symptomatiques et asymptomatiques. Les gouttes infectées par le SARS-CoV-2 peuvent se propager jusqu'à un à deux mètres et peuvent être placées sur des surfaces dans des conditions météorologiques favorables pendant plusieurs jours. Néanmoins, face aux désinfectants courants tels que l'hypochlorite de sodium, le peroxyde d'hydrogène, etc., il est rapidement éliminé. Selon une étude sur d'autres traits du COVID-19, la sensibilité de la maladie sur les personnes dépend de l'âge, de la santé physique et des caractéristiques biologiques. Statistiquement, la plupart des patients adultes ont entre 35 et 55 ans et sont moins fréquents chez les nourrissons et les enfants. Parmi eux, les personnes dont la fonction immunitaire est plus faible, les personnes âgées d'une moyenne d'âge de 60 ans et plus, les personnes souffrant de dysfonctionnement rénal et hépatique et d'hypertension, de diabète, d'asthme, d'obstruction pulmonaire chronique, de patients cardiaques, de fumeurs, de femmes enceintes et de personnes handicapées sont plus à risque et plus susceptibles d'être exposés au virus. Actuellement, il existe un fort potentiel d'épidémies chez les humains ; en d'autres termes, les personnes de tout âge sont infectées par l'infection à coronavirus (9). Chez les personnes âgées, le taux de mortalité moyen était plus élevé avec les maladies que chez les jeunes. Il n'y a eu aucun rapport de femmes enceintes recevant des greffes prénatales ou intra-utérines. Selon l'OMS, les conséquences du non-allaitement et de la séparation entre la mère et l'enfant sont plus importantes que le risque d'infection au COVID-19 chez les nourrissons car l'infection chez les nourrissons est généralement bénigne ou asymptomatique (10). L'OMS recommande que les mères suspectées ou confirmées de COVID-19 soient encouragées à commencer ou à continuer à allaiter. Les mères doivent être informées que les avantages de l'allaitement l'emportent largement sur les risques potentiels de transmission (11). Les indicateurs épidémiologiques importants associés à la COVID-19 sont présentés dans le tableau ci-dessous (12).

1.4 **Traitement**

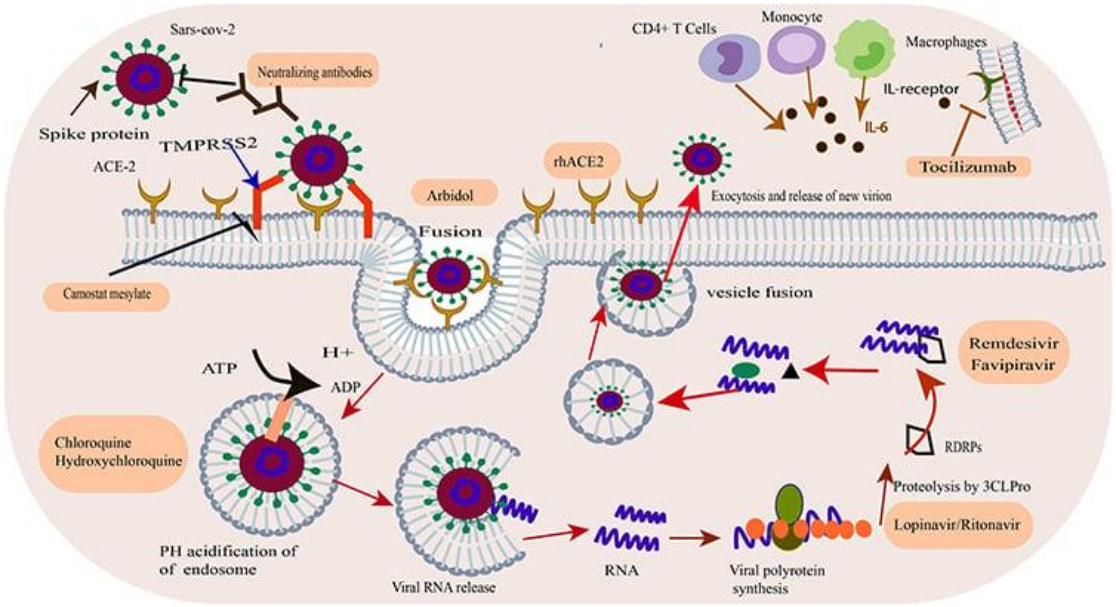


Figure 3: Cycle de réplication du SRAS-CoV-2 avec les différentes molécules et leurs cibles en cours d'évaluation pour traiter la COVID-19 (13).

Le SARS-CoV-2 relie par la protéine S (pour Spike) au récepteur ACE2 exprimé à la surface des cellules cibles, puis pénètre par endocytose ou par fusion directe vers la membrane plasmatique. TMPRSS2 Serine Protease est aussi impliquée dans la phase d'entrée du virus, permettant l'amorçage de la protéine S. Après la libération de l'ARN génomique dans le cytoplasme, il est traduit en polypeptides, qui, clivés par la protéase virale, permettent la génération de protéines non structurales, qui forment le complexe de réplication/transcription (réplicase)*(14)*. L'ARN génomique à polarisation positive est ensuite transcrit en ARN viral à polarisation négative complémentaire, qui sert de matrice pour la synthèse de l'ARN génomique. L'ARN génomique et les protéines structurales vont alors s'assembler dans le réticulum endoplasmique. Les virus néoformés sont alors transportés *via* des vésicules de transport vers l'appareil de Golgi puis vers la surface cellulaire, où ils sont libérés. Les inhibiteurs du SRAS-CoV-2 et leurs cibles connues ou prévues sont représentées par des rectangles à plusieurs stades du cycle viral*(15)*.

Les traitements disponibles pour les personnes atteintes de COVID-19 sont basés sur leurs symptômes, et il n'existe aucun traitement exact disponible pour un rétablissement complet chez les sujets infectés. Les chercheurs et les médecins s'efforcent de fournir un traitement approprié

aux patients COVID19. Les chercheurs testent un large éventail de thérapies possibles, notamment des médicaments antiviraux, des immunosuppresseurs, des anticorps monoclonaux et des vaccins *(16)*. Aux premiers stades de l'infection, le système immunitaire du patient est mis au défi d'empêcher la réplication du virus SARS-CoV-2 ; cependant, dans les stades aigus, il peut subir des lésions tissulaires dues à de graves réactions immunitaires/inflammatoires. Selon la recherche clinique, les thérapies antivirales sont plus efficaces dans les premiers stades de la maladie. En revanche, les traitements immunosuppresseurs/anti-inflammatoires sont susceptibles d'être plus efficaces dans les stades sévères de la COVID-19 *(17)*. Il faut noter que les thérapies à base d'anticorps anti-SRAS CoV-2 sont plus efficaces dans les premiers stades de l'infection avant que le patient n'entre dans la phase aiguë. Par conséquent, les médecins ont recommandé de recevoir des anticorps monoclonaux contre le SARS-CoV-2 *(18)*. Les médicaments approuvés par la Food and Drug Administration (FDA) des États-Unis sont la dexaméthasone et le remdesivir. Il est recommandé pour les patients hospitalisés qui ont besoin d'oxygène supplémentaire *(19)*. Le remdesivir est un médicament nucléotidique intraveineux de l'analogue de l'adénosine. Le mécanisme d'action du Remdesivir contre le virus SARS-CoV-2 est présenté à la Figure ci-dessous qui représente le Remdesivir lié à l'ARN polymérase dépendante de l'ARN et empêche la réplication du virus par arrêt prématuré de la transcription de l'ARN. Dans la phase aiguë de la maladie, lorsque les patients ont besoin d'un ventilateur, la dexaméthasone, un corticostéroïde, affecte de manière significative la récupération des patients *(20)*. Tous les traitements recommandés pour la COVID-19 sont indiqués dans le tableau ci-dessous.

2 Biologie du SARS-Cov-2

2.1 Structure du virus (protéine E, S, M…)

Les coronavirus appartiennent à la sous-famille des Coronavirinae dans la famille des Coronaviridae et la sous-famille contient quatre genres : Alpha, Beta, Gamma et Delta qui possèdent des structures en forme de couronnes vu au microscope électronique (21). La morphologie observée du SARS-CoV-2 est cohérente avec celle des autres membres de sa famille. Le SARS-CoV-2 est une particule enveloppée sphérique contenant de l'ARN simple brin (sens positif) associé à une nucléoprotéine enfermée dans une capside constituée d'une protéine de matrice.et l'enveloppe virale qui porte à sa surface des projections formées de glycoprotéines ce qui donne un aspect en couronne à la particule virale (22).

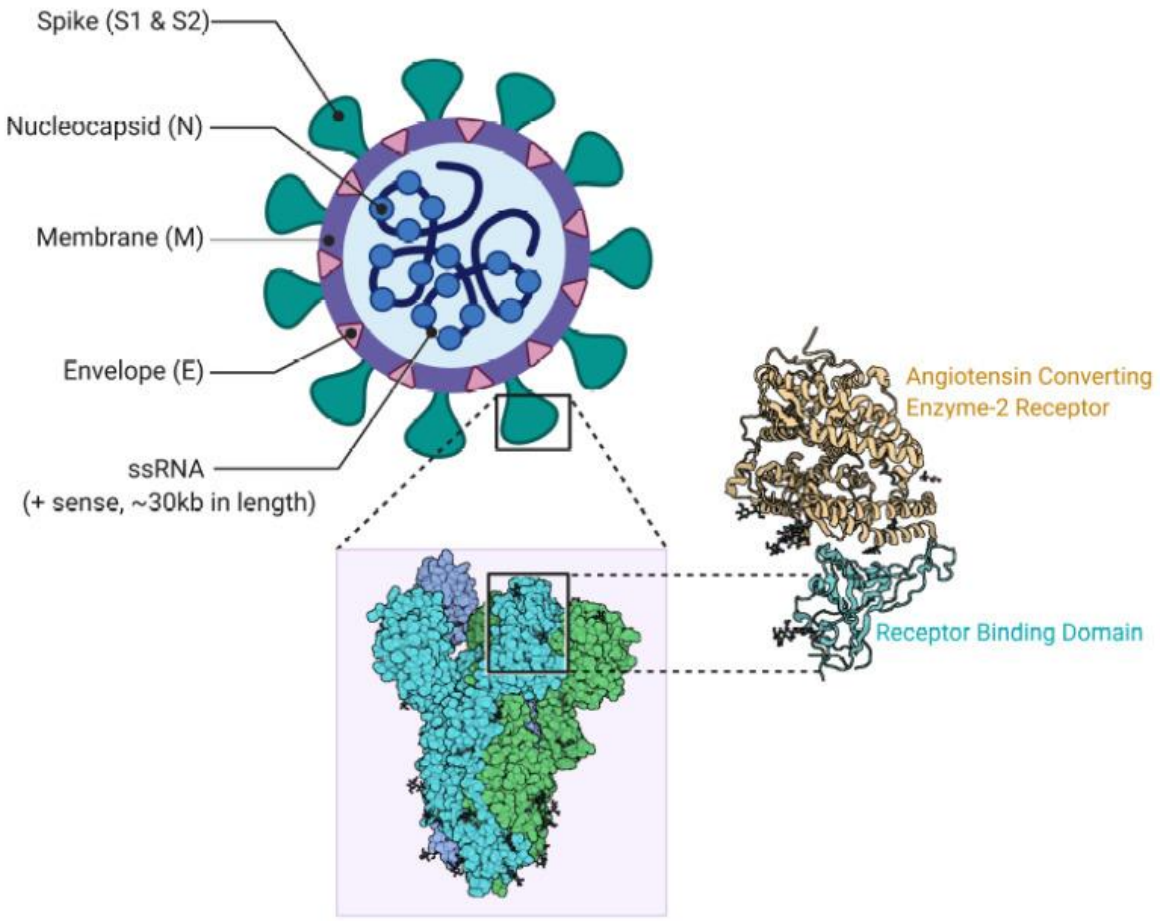


Figure 4: Structure du Sars-COV-2 (6).

Comme montrée sur la figure 5 : Le SARS-CoV-2 possède quatre protéines structurales principales, dont on trouve la protéine de pointe (glycoprotéine de Spike (S)) qui est le médiateur de la liaison au récepteur et de la fusion membranaire, la glycoprotéine de petite Enveloppe (E), la glycoprotéine membranaire (M) et la protéine de la nucléocapside (protéine N) qui est la plus abondante dans les coronavirus, ainsi que plusieurs protéines accessoires (23).

2.1.3 Protéines structurales

A. La protéine Spike(S)

Le SARS-CoV-2 possède à sa surface une protéine nommée Spike ou protéine S ou protéine de pointe. Cette protéine transmembranaire de type I est présente sur l'enveloppe virale sous la forme d'un trimère. Chaque monomère se compose de 2 sous-unités, S1 et S2. La sous-unité S1 contient un domaine RBD (receptor binding domain) qui se fixe au récepteur ACE2 sur la membrane de la cellule hôte. La sous-unité S2 est activée par la liaison à S1 et le récepteur ACE 2 et contient les éléments nécessaires à la fusion membranaire (22). La protéine S joue donc un

rôle important dans l'infection par le SARS-CoV-2 ainsi que dans l'induction des réponses des anticorps de neutralisation et des cellules T et dans l'immunité protectrice (24).

B. La protéine d'enveloppe (E)

La protéine d'enveloppe, ou protéine E, est un petit polypeptide, qui est impliqué dans l'assemblage, le bourgeonnement , la formation de l'enveloppe et la pathogénie tels que l'assemblage, et qui assure également le fonctionnement des canaux ioniques de la protéine E et la modification de l'équilibre ionique des cellules de coronavirus qui est un processus nécessaire pour la production de virus , elle joue également un rôle dans l'induction de l'apoptose, active l'inflammasome NLRP3 de l'hôte, ce qui entraîne une surproduction d'IL-1 bêta (21).

C. La protéine de membrane (M)

La protéine M traverse trois fois la bicouche membranaire, laissant un court domaine N-terminal en dehors du virus et un long domaine C-terminal à l'intérieur du virion. Ce dernier peut interagir avec la protéine de la nucléocapside (protéine N). La protéine M est impliquée dans la formation intracellulaire des particules virales. Etant donné que la protéine M est la protéine la plus abondante à la surface du virus, des anticorps contre cette protéine sont retrouvés dans le sérum des patients. Donc la protéine M est un outil primordial en diagnostic (24).

D. La protéine de nucléocapside (N)

La protéine de nucléocapside qui est également intitulée de ribonucléoprotéine qui se présente en abondance Cette phosphoprotéine de 50 kDa codée par le gène ORF9b est impliquée dans la réplication du génome viral et dans la modulation des voies de signalisation cellulaire. Lors de l'assemblage du virion, la protéine N se lie à l'ARN viral ce qui aboutit à la formation de la nucléocapside hélicoïdale. En raison de la conservation de la séquence de la protéine N et de sa forte immunogénicité, la protéine de nucléocapside du coronavirus est choisie comme outil de diagnostic ou comme cible potentielle pour la mise au point de nouveaux vaccins (25).

Figure 5: Organisation génomique du Sars-CoV-2 (23).

2.1.3 Protéines non-structurales (NSP)

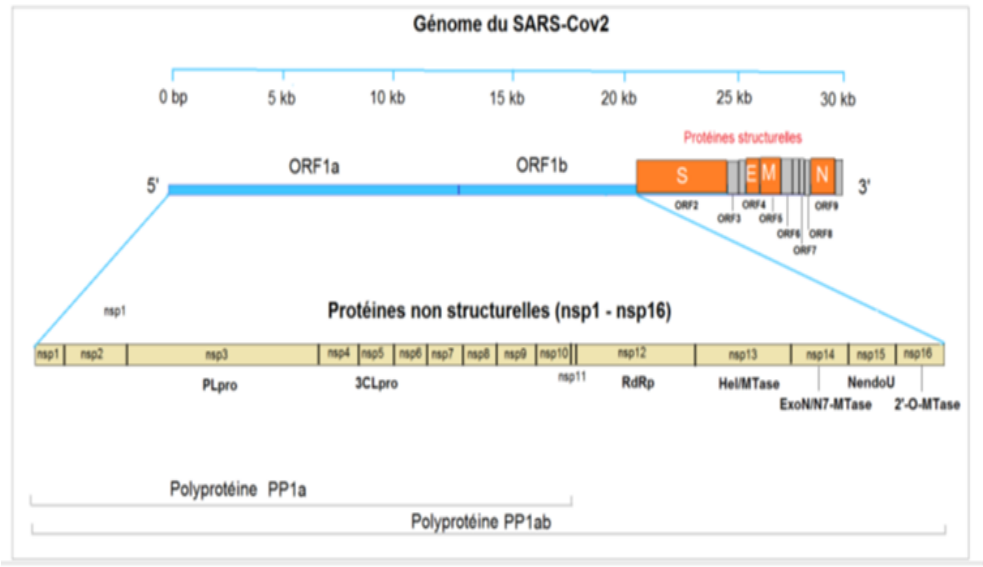


Figure 6: Les protéines non -structurales (NSP)(26).

Le génome du SARS-CoV-2 comporte un nombre variable de cadres de lecture ouverts (ORF). Les deux niveaux de l'ARN viral sont situés principalement dans le premier ORF (ORF 1 a/b), traduit en deux polyprotéines, pp1a et pp1b et code 16 protéines non structurales (nsp 1 à nsp 16) .de nsp1 à nsp11 sont codés par ORF1a et nsp 12 à nsp 16 sont codés par ORF1b, alors que les ORF restants codent pour les 4 protéines structurales et les 9 protéines accessoires. Les NSP comprennent les divers enzymes et facteurs de transcription que le virus utilise pour se répliquer, tels que la protéase virale, la réplicase de l'ARN et les protéines de contrôle de l'hôte (27).

2.1.3 Protéines accessoires

Le SARS-CoV-2 possède également des protéines accessoires dérivées de l'ARN génomique et réparties entre les gènes structuraux : ORF (3a, 3b, 6, 7a, 7b, 8b, 9b et 14). Qui sont impliquées dans la pathogénicité virale en modulant par exemple les voies de signalisation de l'interféron de l'hôte (28).

- ORF3a : La protéine est primordiale pour la pathogenèse de la maladie. Elle contient six domaines fonctionnels (I à VI) dont la protéine est liée à la virulence, à l'ineffectivité, à la formation de canaux ioniques et à la libération de virus.

- ORF3b : L'une de ses particularités est la présence de codon stop prématuré dans son gène ORF3b. qui se considère un antagoniste puissant de l'interféron (IFN) en supprimant son induction.
- ORF6 : c'est la protéine qui assure un rôle essentiel dans la pathogenèse virale.
- ORF7A : protéine transmembranaire de type I.
- ORF9 : protéine structurelle qui se lie directement à l'ARN viral et assure sa stabilité.
- ORF10 : C'est une protéine hypothétique de 38 acides aminés qui ne partage aucune similitude de séquence avec aucune autre protéine connue (29).

2.2 **Cycle de vie**

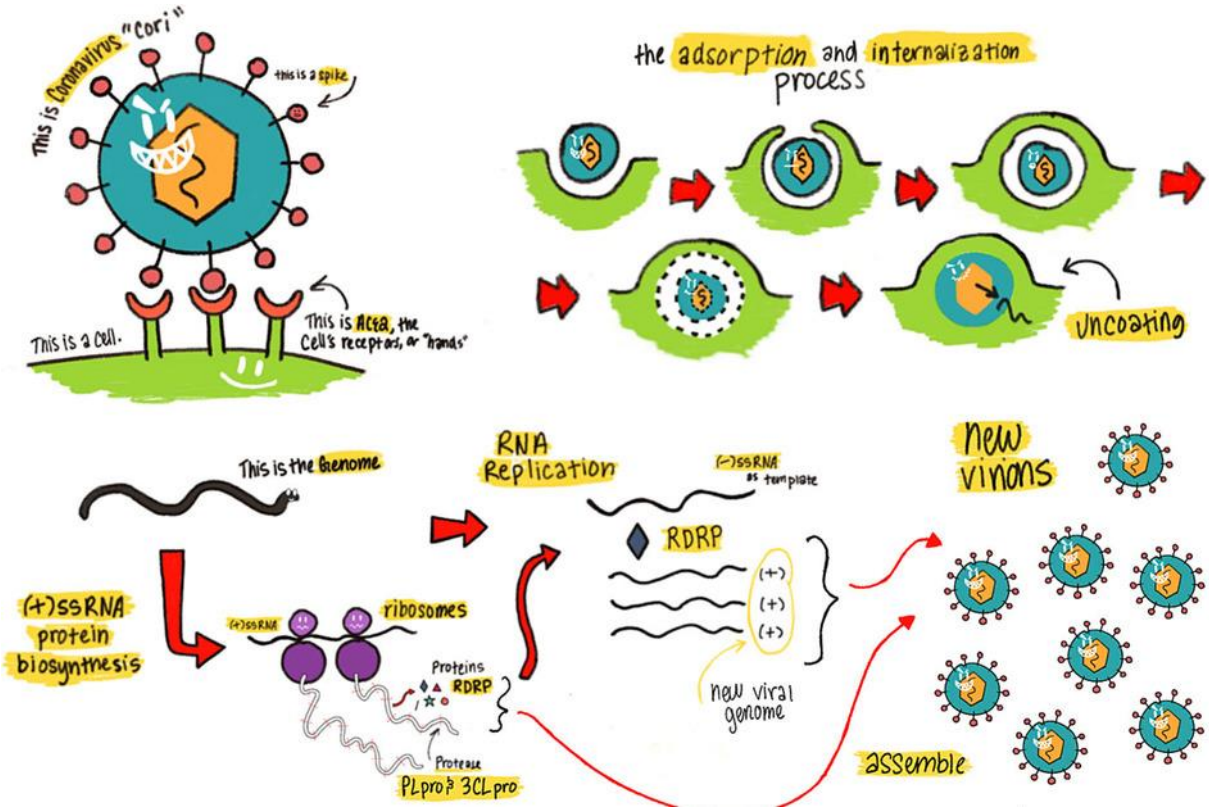


Figure 7: Les étapes du cycle viral du SARS-CoV-2 et les cibles thérapeutiques. TMPSS2 : protéase transmembranaire à sérine 2. ACE2 : enzyme de conversion de l'angiotensine 2 (30).

Le cycle du virus dans la cellule se décompose en trois grandes étapes : l'entrée du virus dans la cellule hôte, la réplication du génome et la formation et la sécrétion de nouveaux virions (31). Le SARS-COV-2 commence son cycle lorsque sa protéine S se lie au récepteur ACE2 exprimé à la surface des cellules-cibles et pénètre ensuite par endocytose ou par fusion directe à la membrane plasmique. La sérine protéase TMPRSS2 est également impliquée dans l'étape d'entrée du virus, en permettant l'amorçage de la protéine S, ensuite le virus libère son ARN

dans la cellule hôte pour être traduit en protéines qui sont secondairement clivées pour former les protéines structurales (32).

L'ARN génomique et les protéines structurales sont ensuite assemblées en virions dans le réticulum endoplasmique et transportées via des vésicules de transport vers l'appareil de Golgi puis vers la surface cellulaire. Les virus néoformés sont alors transportés via des vésicules de transport vers l'appareil de Golgi puis vers la surface cellulaire où ils sont libérés (33).

3 Outils de diagnostic biologique.

En raison de l'absence de traitement curatif définitif pour cette maladie, la solution la plus efficace après la prévention et le contrôle est le diagnostic rapide de la maladie et l'isolement des malades. Il existe plusieurs façons de diagnostiquer la maladie de manière précoce, telles que la méthode RT-PCR, le CT- Scan, le test sanguin sérologique d'anticorps et l'intelligence artificielle (34).

3.1 Méthode RT-PCR

L'un des moyens les plus importants de détecter le virus SARS-CoV-2 dans les échantillons des voies respiratoires supérieures et inférieures est le panel de diagnostic PCR en temps réel. La base de la PCR consiste à copier la structure de l'ARN et de l'ADN de l'échantillon, ce qui permet de diagnostiquer l'origine infectieuse et diverses maladies génétiques et sanguines. La figure ci- dessous illustre les tests de diagnostic COVID-19 par RT-PCR en temps réel (29,30). Comme le montre cette figure, il y a cinq étapes nécessaires pour effectuer le test : prélèvement d'échantillons, extraction d'ARN, configuration RT-qPCR et ensuite l'affichage des résultats des tests, qui peuvent tous être personnalisés pour expliquer à la fois cela et d'autres protocoles de diagnostic RT-qPCR.

Les résultats des tests des patients positifs au SARS-CoV2 franchissent la ligne de seuil en 40,00 cycles. Cette méthode permet de détecter l'acide nucléique présent dans le prélèvement d'un écouvillon nasal ou dans les voies respiratoires à l'aide du PCR en temps réel. Elle est confirmée sur la base de la fonction de reproduction et de la séquence du virus dans l'échantillon (37).

Étant donné que le virus infectieux infecte le système respiratoire de l'hôte, les échantillons nécessaires sont prélevés dans les voies respiratoires inférieures et supérieures. Le test sur écouvillon consiste à prélever un échantillon spécial dans la gorge et le nez d'une personne. La

RT-PCR est largement utilisée dans le domaine du diagnostic et le pourcentage d'erreur de la méthode est faible (38).

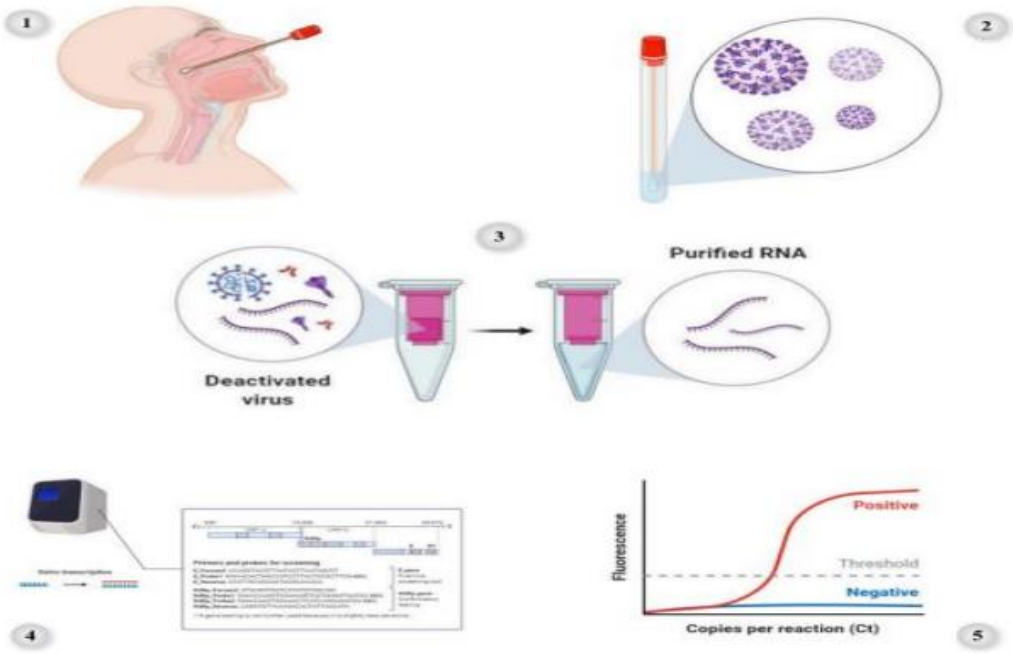


Figure 8: Le modèle de protocole test de diagnostic COVID-19 par RT-PCR en temps réel (39).

3.2 **Tomodensitométrie :**

La tomodensitométrie (CT) est une méthode de diagnostic afin d'examiner les schémas morphologiques des lésions pulmonaires associées au COVID-19 consiste à utiliser des scanners thoraciques tels que les rayons X et les tomodensitogrammes (TDM). Il convient de noter que la précision du diagnostic dépend fortement des spécialistes. Aux premiers stades de l'épidémie, dans n'importe quel pays, l'utilisation de méthodes d'imagerie CT était plus critique que la RT-PCR en raison du manque de technologie RT-PCR ou du manque de kits et d'équipements de diagnostic adaptés à un échantillonnage précis (40). Ainsi que les images CT aident les médecins à identifier les structures internes et de déterminer leur forme, leur taille, et leur densité. autrement dit l'imagerie CT aide à mettre en évidence les anomalies causées par le COVID-19 (41). Chez environ 85 % des patients présentant des lignes et des interfaces irrégulières superposées, le scanner thoracique dans les cas de pneumonie COVID-19 montre des opacités en verre dépoli (GGO) bilatérales, périphériques et basales prédominantes et/ou une consolidation. De plus, dans la patience sans détresse respiratoire sévère qui s'est remise de la maladie à coronavirus 2019, les tomodensitogrammes thoraciques ont montré que la plus grande gravité des anomalies pulmonaires s'est produite environ dix jours après les premiers

symptômes (42). Il est donc primordial de porter une attention particulière à la patience asymptomatique car ils peuvent transmettre l'infection . L'imagerie CT de ces patients présente des caractéristiques uniques qui peuvent être détectées même en cas de patience asymptomatique avec des tests nucléiques négatifs. La figure ci-dessous montre les résultats positifs et négatifs de la tomodensitométrie COVID-19 du patient (43).

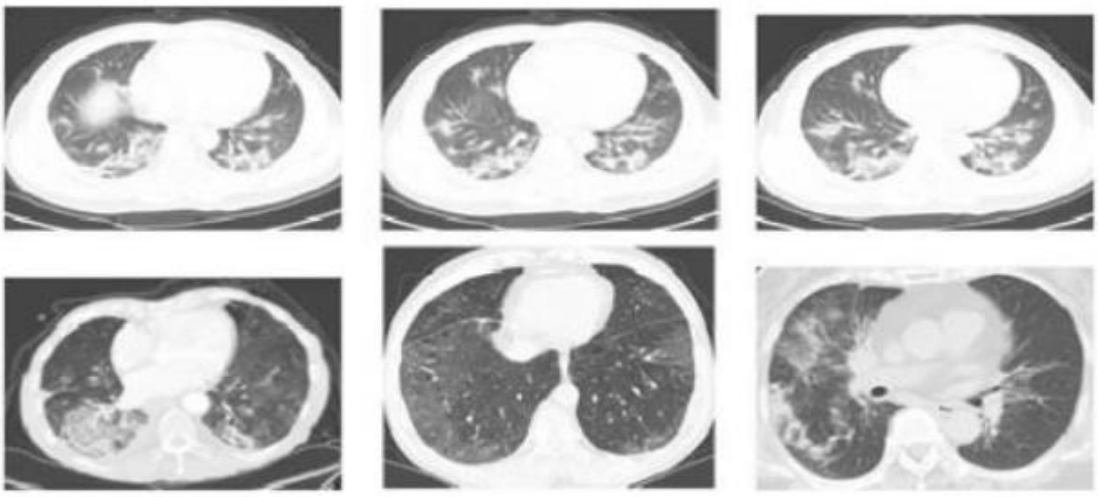


Figure 9: Exemples d'images CT positives pour COVID-19 (en haut) et non COVID-19 (en bas) à partir de l'ensemble de données COVID-CT (44).

3.3 Le test sanguin sérologique d'anticorps :

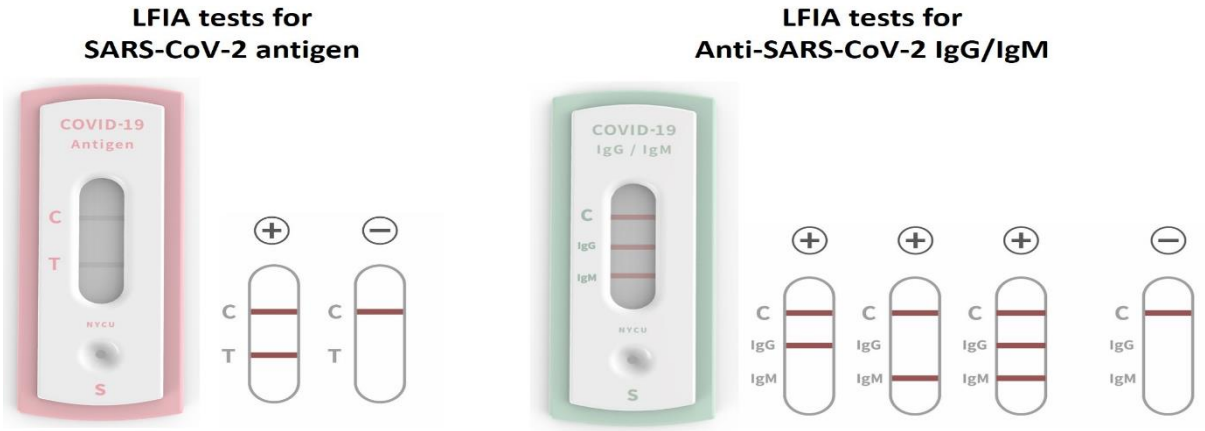


Figure 10: représentation des dispositifs détectables des antigènes et des anticorps IgG et IgM (45).

Le test sérologique est une méthode de diagnostic qui permet de détecter les réponses immunitaires médiées par les anticorps contre les agents infectieux. Le Centre européen de contrôle et de prévention des maladies (ECDC) a approuvé le test sérologique COVID-19 à des fins épidémiologiques et de surveillance uniquement parce qu'il ne détecte pas les premiers

stades de l'infection. Les tests sérologiques rapides présentent une alternative aux tests moléculaires pour identifier les patients COVID-19 lorsque l'accès aux tests PCR est limité ou inexistant (23). L'utilisation de tests sérologiques à faible prévalence n'est pas appropriée car cette méthode est susceptible d'avoir des résultats faussement positifs par rapport aux vrais positifs. Ce modèle de protocole représenté dans la figure illustre les tests de diagnostic sérologique COVID-19 par détection d'anticorps. Il décrit le chargement de l'échantillon, la détection de l'anticorps-antigène du SRAS-CoV-2 et les résultats des tests qualitatifs (46). Cela peut être personnalisé dans son ensemble pour expliquer d'autres protocoles de diagnostic sérologique pour différents agents pathogènes viraux, bactériens ou parasitaires (47).

II Outils de séquençage

Le génome du nouveau syndrome respiratoire aigu sévère 2 (SARS-CoV-2) a une taille comprise entre 29,8 et 29,9 kb, et sa séquence diffère de certains virus corona humains précédemment identifiés, notamment le Syndrome respiratoire oriental (MERS). Cependant, une enquête appropriée sur les caractéristiques épidémiologiques, virologiques et pathogènes du SARS-CoV-2 est essentiel pour introduire de nouvelles approches de traitement et développer des stratégies de prévention efficaces. Pour les outils et techniques bio-informatiques, on va les mettre en œuvre (48).

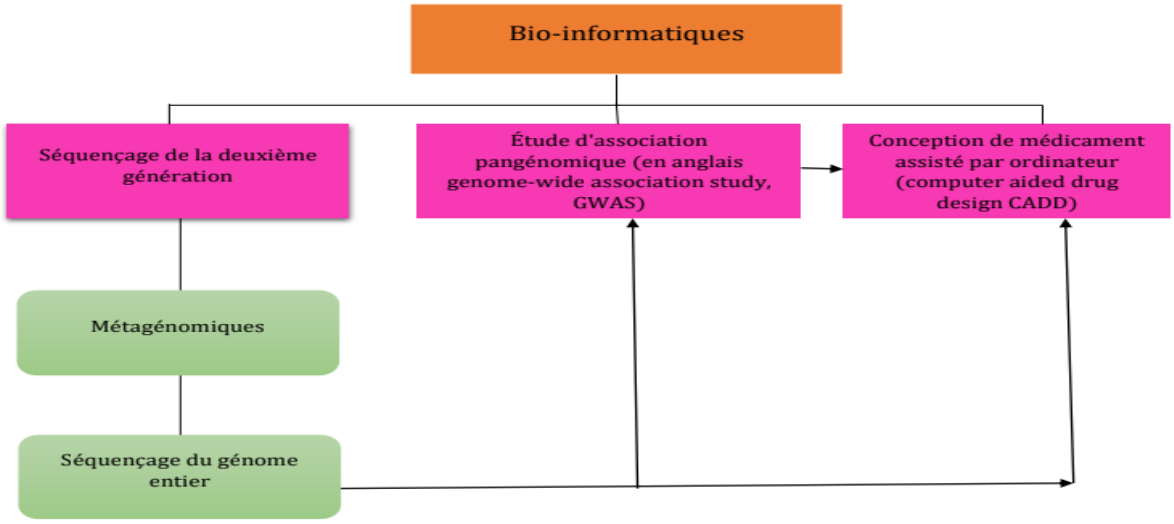


Figure 11: La représentation graphique des applications bio-informatiques interconnectées mises en œuvre dans la recherche COVID-19 (49).

Diverses techniques de séquençage peuvent être utilisées. Permettant de cibler l'analyse génétique lorsque l'on a déjà des informations assez précises sur le gène en cause, ainsi de rechercher des anomalies dans plusieurs gènes à la fois (panel de gènes) lorsque l'information n'est pas assez précise (50).

1 Séquençage de la deuxième génération :

Le progrès dans les innovations de séquençage de nouvelle génération (NGS) a entraîné une multiplication remarquable des données de séquence génomique. NGS a métamorphosé la profondeur des sciences biomédicales. Lors d'une épidémie dans un système de soins de santé, l'identification rapide et efficace de l'agent pathogène causatif avec des enquêtes épidémiologiques est indispensable pour permettre une réaction ciblée sur le contrôle de la maladie. La précision du NGS dans les variantes virales a analysé et quantifié de manière productive la diversité extrêmement élevée au sein des quasi-espèces virales. De nombreuses mutations résistantes aux médicaments ou aux vaccins ont été découvertes à faible fréquence et reflètent une importance thérapeutique. Les technologies de séquençage à haut débit, y compris la technique de méta génomique de séquençage du génome entier (WGS) ce qui permet l'obtention rapide de la séquence complète des génomes pathogènes (3).

1.1 Séquençage illumina :

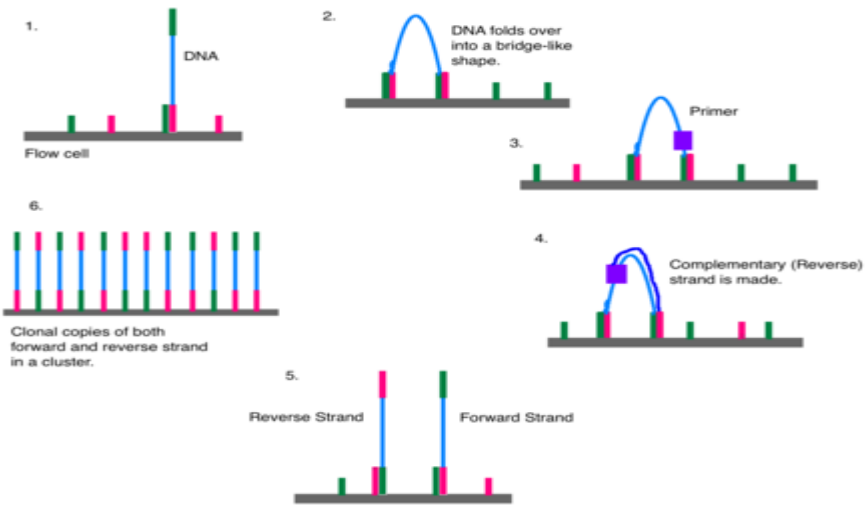


Figure 12: Étapes de préparation de la bibliothèque pour le séquençage Illumina (51).

Illumina est une technique qui permet le séquençage de plusieurs centaines de millions de fragments d'ADN. Son application contribue l'étude de la variété microbienne dans l'environnement, de séquencer des génomes entiers, ou d'analyser les interactions entre ADN et protéines. Les domaines d'applications sont les suivants : la recherche sur le cancer, les analyses criminalistiques, la FIV (Fécondation In Vitro), le DPI (Diagnostic Préimplantatoire), l'agri génomique, la découverte de virus ou l'identification de micro-organismes non cultivables (52).

Le fonctionnement de cette technologie se présente par la disposition des fragments d'ADN sur une plaque recouverte de deux types d'oligonucléotides chacun identiques à l'une des extrémités des brins d'ADN. L'ADN complémentaire à ces fragments est synthétisé en se servant des oligonucléotides de la plaque comme support initiateur. Les fragments primaires d'ADN sont dénaturés puis lavés à l'aide d'un tampon spécifique. Le brin se plie et s'hybride au second type d'oligonucléotides présent sur la plaque. Puis on assiste à la synthèse du brin anti sens en formant un pont double brin. Sa dénaturation entraîne la formation d'un brin linéaire sens et un brin linéaire anti sens identique au brin d'origine.

Cette étape est répétée plusieurs fois. Après l'amplification, les ponts sont coupés et la partie inversée et lavée grâce à un tampon. Afin de garantir la synthèse du brin dans le bon sens, les trois premiers nucléotides attachés aux oligonucléotides du tapis sont « bloqués ». L'ajout de l'amorce de séquençage catalyse la polymérisation réalisée à l'aide des nucléotides marqués avec une sonde fluorescente. Les clusters sont excités à une longueur d'onde donnée entraînant ainsi l'émission d'un signal lumineux (ce type de séquençage est appelé « séquençage par synthèse »). Le nombre de cycles détermine la longueur du fragment, la longueur d'onde et l'intensité du signal émis déterminent le nucléotide associé. Tous les brins identiques sont simultanément lus. Une fois que la première lecture est achevée le brin dernièrement synthétisé est éliminé.

L'introduction et l'hybridation du premier index au brin suivie par synthèse jusqu'à atteindre les oligonucléotides. Une fois complété, l'index 1 est lavé et les trois derniers nucléotides sont déprotégés. Le brin se plie alors et se lie au deuxième type d'oligonucléotide. L'index 2 est synthétisé de la même manière afin d'assurer la fixation de l'ADN polymérase qui allonge l'oligonucléotide en formant un pont double-brin qui sera dénaturé, engendrant deux brins linéaires. Les extrémités 3' sont bloquées et le brin sens est éliminé laissant seulement le brin anti sens.

1.2 **Séquençage Ion Torrent :**

Le séquenceur Ion Torrent combine la technologie des semi-conducteurs préexistants avec une chimie de séquençage simple Le séquenceur est essentiellement un pH-mètre à l'état solide (53).Reconnue par son extensibilité, sa simplicité et sa rapidité, cette technologie peu coûteuse offre une gamme d'options allant de 500 000 lectures sur une puce 314 à 5 millions de lectures et sur une puce 318 ainsi qu'une longueur de lecture de 200 et 400b. Les exécutions de séquençage ne durent que 4 à 8 heures par rapport à quelques jours d'exécution sur Illumina. Un seul échantillon est exécuté à la fois. Il n'y a donc pas d'attente nécessaire pour remplir, une Flow Cell, une plaque ou une lame contrairement aux autres plateformes de séquençage de nouvelle génération (54).

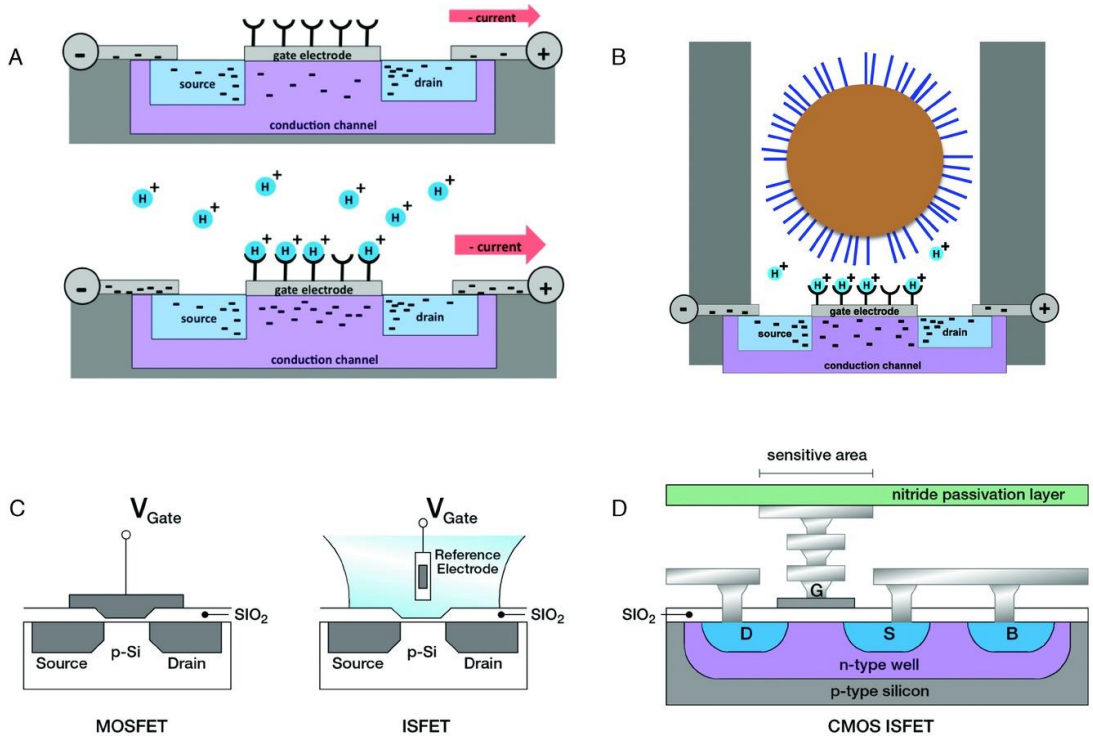


Figure 13: Les différentes étapes du séquençage avec le système ion torrent (55).

Le système ion torrent utilise un semi-conducteur à oxyde métallique complémentaire (CMOS) et le système utilise également un transistor à effet de champ sensible aux ions (ISFET) ce qui permet de détecter les ions H+ libérés lors de l'incorporation du nucléotide

Lors de l'incorporation des nucléotides il y'a la libération d'un ion H+ ce qui aboutit au changement du PH de la solution qui est détecté grâce au capteur CMOS.

Le séquençage Ion Torrent est constitué de 4 étapes principales :

- **La construction de banque** : Il s'agit de fragmenter l'ADN en fragments de taille uniforme (généralement 200-400 Pb) puis l'ajout des adaptateurs de séquençage.

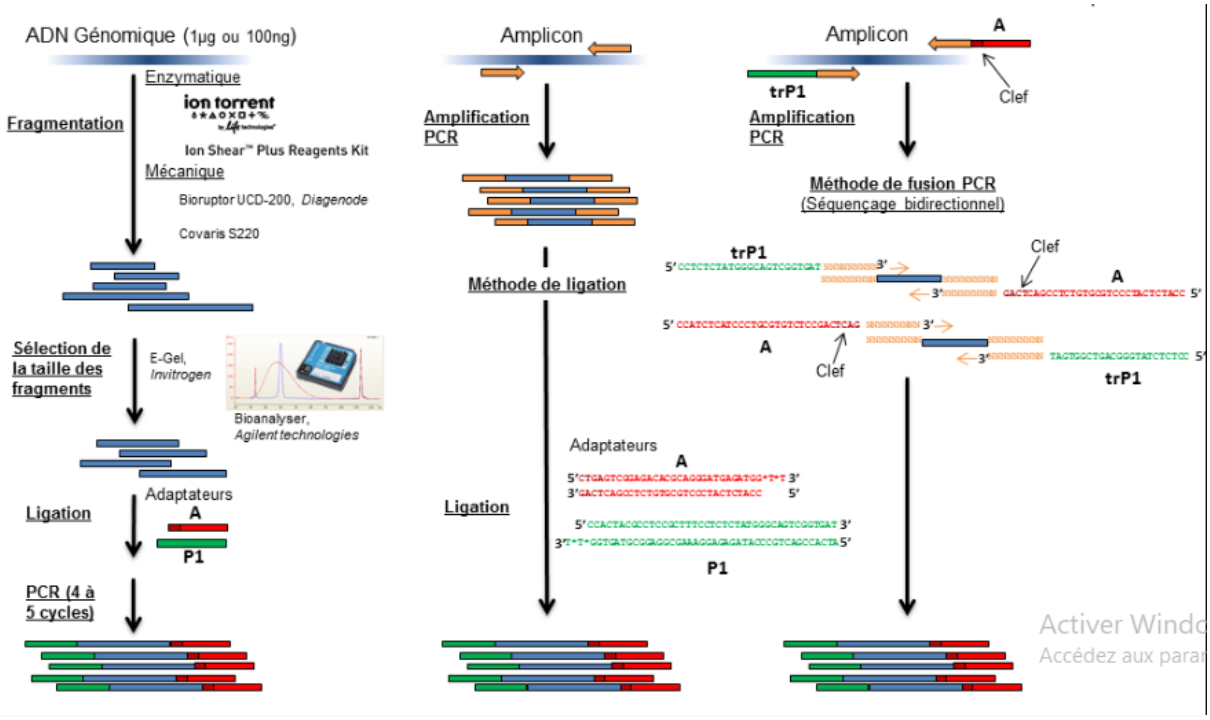


Figure 14: Préparation de la librairie (56).

- **L'amplification**

Les fragments produits lors de la préparation de la banque sont attachés à des billes et puis amplifiés par PCR en émulsion, des billes revêtues d'amorces complémentaires sont mélangées avec une solution aqueuse diluée contenant les fragments à séquencer avec les réactifs PCR nécessaires ,ensuite la solution est mélangée avec de l'huile pour former une émulsion de microgouttelettes, la concentration de billes et de fragments est maintenue suffisamment basse pour que chaque microgouttelettes ne contienne qu'une de chaque, une amplification clonale de chaque fragment est ensuite réalisée au sein des microgouttelettes, après amplification l' émulsion est rompue par extraction organique et centrifugation, les billes amplifiées sont enrichies dans un gradient de glycérol avec des billes non amplifiées se granulant au fond donc le processus de PCR en émulsion est efficace lent et compliqué (57).

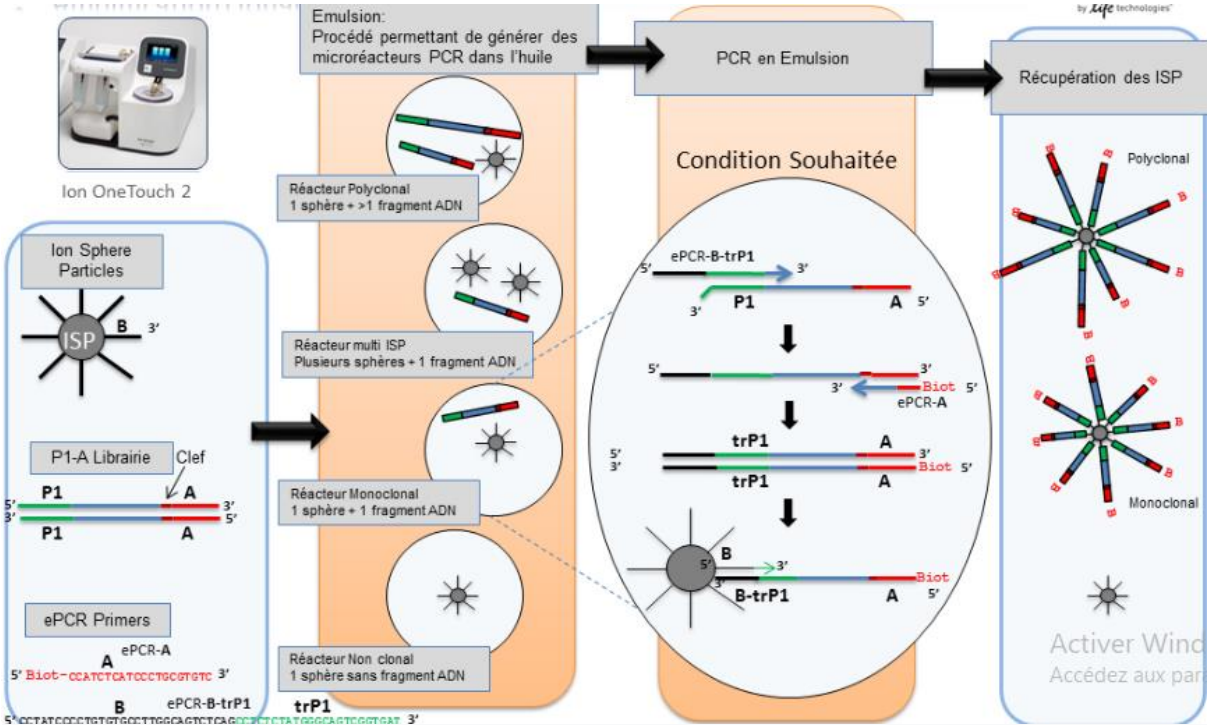


Figure 15: Préparation de la matrice de séquençage (58).

- **<u>Le séquençage :</u>**

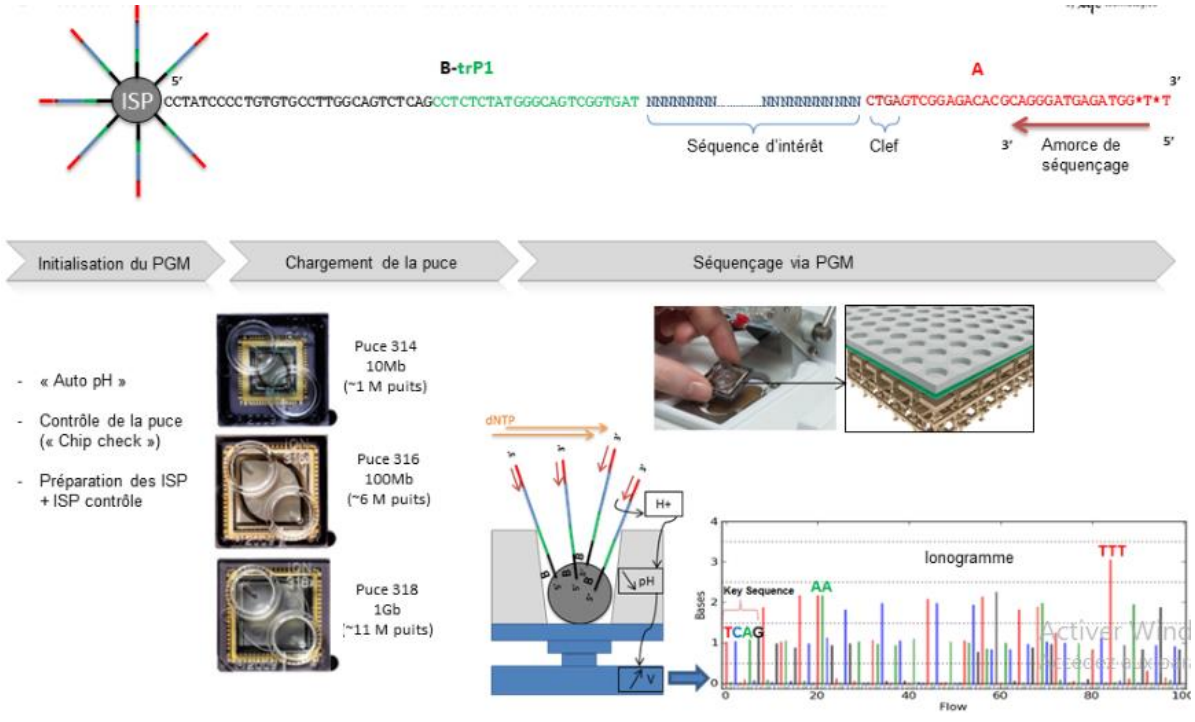


Figure 16: Le principe du séquençage par ion torrent PGM (57).

Basé sur la chimie standard du pyroséquençage qui correspond à la forme de « séquençage par synthèse » les bases individuelles sont introduites toutes à la fois à l'aide par l'ADN polymérase, le système Ion torrent mesure la libération directe de H+ à partir de la réaction, des instruments relativement peu coûteux couplés à puces jetables, qui agissent essentiellement comme des pH-mètres, le système peut utiliser des nucléotides non modifiés, qui sont moins chers et mieux tolérés par l'ADN polymérase, chaque perle est placée dans un seul puits d'une lame, comme 454, la lame est inondée d'une seule espèce de dNTP, ainsi que des tampons et de la polymérase et en même temps un NTP ensuite la détection du Ph dans chaque puits ,comme chaque ion H+ libéré diminue le ph ce qui permet de déterminer si cette base et le nombre de bases qui a été ajouté à la séquence lue ensuite les dNTP sont lavés et le processus se répète en passant par les différentes dNTPs (57).

- **L'analyse :**

Consiste à générer des fichiers de sortie standard comme FASTQ, l'analyse des données est généralement simple, ion torrent propose le navigateur torrent, il s'agit d'un logiciel qui agit comme l'interface principale pour un certain nombre de fonctions de base (59).

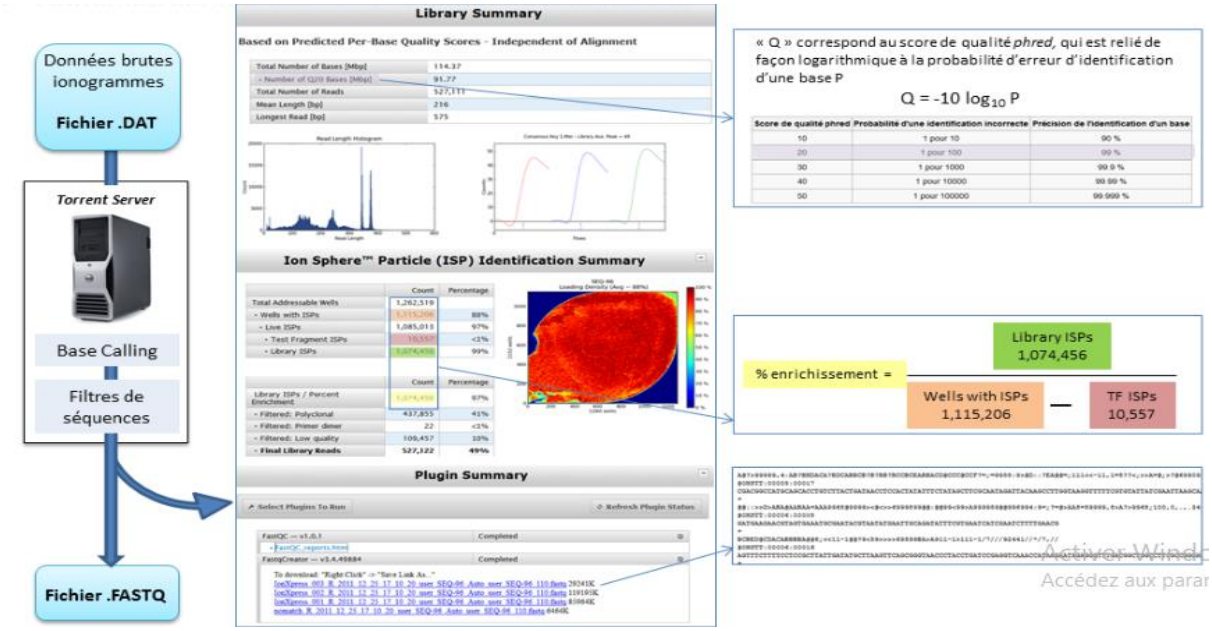


Figure 17: Données de séquençage (58).

Contrairement au séquençage Illumina, le séquençage Ion Torrent ou Ion proton n'utilise pas de signaux optiques. Il utilise 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 Pb. 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.

Matériels et méthodes

I Collecte des échantillons

1 Séquenceur Illumina MiSeq

Figure 18: Séquenceur de nouvelle génération Illumina MiSeq(60).

Les échantillons ont été obtenus à partir de prélèvements naso-pharyngés sur des patients ayants COVID-19. L'ARN viral a été directement extrait à partir du prélèvement en utilisant le kit QIAamp viral RNA mini kit (Qiagen, Germany). L'ADNc et la préparation de librairie pour le séquençage NGS a été effectué en utilisant le kit QIAseq FX DNA library Qiagen, Germany). Les échantillons préparés ont été séquencés en paired-end en utilisant Illumina MiSeq.

La séquence de référence du syndrome respiratoire aigu sévère coronavirus 2 Wuhan-Hu-1 a été collectée depuis GenBank dans le National Center for Biotechnology Information (NCBI) (numéro d'accession de cette séquence : NC_045512.2).

2 Séquenceur Ion Torrent

Figure 19: Séquenceur ion Torrent(61).

Le prélèvement a été effectué le 27 décembre 2021. L'ARN a été extrait d'un prélèvement nasopharyngé d'un patient de la Région de Rabat-Salé-Kenitra, Maroc, à l'Hôpital Ibn Sina de Rabat, à l'aide du kit de purification d'ADN/ARN du virus MegaPure en le système de purification d'acide nucléique-32 (Bigfish Bio-tech, Hangzhou, Chine). Le patient a été identifié comme positif pour COVID-19 par PCR quantitative de transcriptase inverse à l'aide d'un kit SARS-CoV-2 (MAScIR, Maroc) et présentait des valeurs de seuil de cycle (CT) de 18 et 19 pour les gènes RdRp et S, respectivement. L'ADNc total a été préparé à l'aide d'un kit de synthèse d'ADNc SuperScript VILO (Invitrogen, Thermo Fisher Scientific, USA) et utilisé pour la préparation de la librairie SARS-CoV-2 avec un kit Ion AmpliSeq pour Chef DL8 (Thermo Fisher Scientific). La bibliothèque a été ajustée à 30 picomoles (pM) et chargée sur l'instrument Ion Chef (Thermo Fisher Scientific) pour la PCR en émulsion, l'enrichissement et le chargement sur la puce Ion S5 530. Le séquençage du génome entier (WGS) a été réalisé à l'aide du panel de recherche Ion AmpliSeq SARS-CoV-2 (Invitrogen, Thermo Fisher Scientific) pour un séquençage complet du génome viral conformément aux instructions d'utilisation sur un système de la série Ion GeneStudio S5 Prime.

II Analyse génomique

Le génome du SARS-CoV-2 étant tout petit (30 kbases) par rapport au génome humain (3 Gbases), l'analyse bio-informatique peut se faire avec un ordinateur personnel en se basant sur les outils d'alignement et de détection des variantes.

Dans notre étude, nous avons travaillé avec 2 séquences issues de deux différents séquenceurs. Cela dit le workflow d'analyse génomique diffère légèrement entre les 2. Nous allons commencer par le workflow d'analyse d'un génome de SARS-CoV-2 à partir des données brutes générées par le séquenceur Illumina. (Annexe Figure 1)

Pour cela, on a récupéré les données de séquençage contenant des courtes séquences, appelées reads, générées par le séquenceur. On les ensuite toutes alignées sur le génome de référence de Wuhan pour en extraire les mutations qu'on annotées. Enfin, on a reconstruit le génome de l'échantillon afin de pouvoir attribuer la lignée.

Cette analyse exige un terminal Linux avec différents outils installés avec conda :

1 Alignement des reads sur le génome de référence de Wuhan

L'alignement vise à aligner les reads présents dans les fichiers Fastq, sur un génome de référence du Sars-CoV-2 qui fait lui environ 30 000 bases. Nous devons d'abord télécharger ce génome de référence surnommé de NC_045512.2. Il s'agit d'un fichier Fasta récupéré via la commande suivante :

wget -O reference. fasta https://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?tool=portal&save=file&log$=seqview&db=nuccore&report=fasta&id=1798174254&extrafeat=null&conwithfeat=on&hide-cdd=on

Avant de procéder à l'alignement avec l'outil bwa, il est indispensable d'indexer ce génome (62).

Bwa index reference. fasta

L'alignement est conçu à l'aide de la commande ci-dessous qui nous créera un nouveau fichier *SAM* contenant les reads associés à leur position d'alignement sur le génome :

bwa mem reference. fasta (R1). fastq (R2). fastq > (ech1). sam

2 La conversion du SAM en BAM

Le résultat de l'alignement de BWA ou d'autres aligneurs est un fichier SAM basée sur le format texte. On doit d'abord convertir le SAM en son homologue binaire sous le format BAM qui est beaucoup plus facile à utiliser pour les programmes informatiques, sauf qu'il est très difficile d'être lu par les humains. Alors pour convertir SAM en BAM, on utilise la commande samtools view (63).

On spécifie que notre entrée est au format SAM par défaut en utilisant l'option -S. Nous devons également dire que nous voulons que la sortie soit BAM par défaut avec l'option -b. Samtools suit la convention UNIX d'envoi de sa sortie vers UNIX STDOUT, nous devons donc utiliser un opérateur de redirection (">") pour créer un fichier BAM à partir de la sortie donc on obtient la commande suivante :

samtools view -S -b ech1.sam > ech1.bam

3 Trier le fichier BAM

Lorsque vous alignez des fichiers FASTQ avec tous les aligneurs de séquence actuels, les alignements produits sont dans un ordre aléatoire par rapport à leur position dans le génome de référence. En d'autres termes, le fichier BAM est dans l'ordre dans lequel les séquences se sont produites dans les fichiers FASTQ d'entrée. Alors pour appeler des variantes ou visualiser des alignements, il faut que le BAM soit davantage manipulé et trié de manière à ce que les alignements se produisent dans "l'ordre du génome". C'est-à-dire classés par position en fonction de leurs coordonnées d'alignement sur chaque chromosome

Samtools sort (ech1). bam -o (ech1.bam_sorted). bam

4 Indexation du fichier BAM

L'indexation d'un fichier BAM trié par génome permet d'extraire rapidement des alignements chevauchant des régions génomiques particulières. De plus, l'indexation permet de facilité la recherche des régions génomiques.

Samtools index (ech1.bam_sorted). bam

5 Appel des variations (variant calling)

Variant calling est le processus par lequel on identifie des variantes à partir de données de séquençage et donc qui permet d'identifier où les lectures alignées diffèrent du génome de référence et de les marquer dans un fichier VCF (64).

Bcftools mpileup –f reference. fasta (ech1). bam | bcftools call –ploidy 1 -mv -o (seq).vcf

La première partie mpileup génère des probabilités de génotype à chaque position génomique avec couverture. La deuxième partie d'appel effectue les appels réels. Le commutateur -m indique au programme d'utiliser la méthode d'appel par défaut, l'option -v demande de n'afficher que les variantes de sites, enfin l'option -o sélectionne le format de sortie.

6 Annotation des variantes

L'annotation des variantes avec SnpEff est considérée une étape primordiale qui permet l'obtention des informations sur la classe fonctionnelle et les effets prédictifs du variant. Tels que la révélation sur la protéine codée, de prédire le rôle de la région sélectionnée ainsi que la localisation de la région d'intérêt au sein d'un génome (65).

SnpEff est un outil utilisé pour annoter nos variantes basées sur Java, similaire à picard. Pour exécuter la commande snpEff, il faut spécifier deux choses :

- Le génome approprié
- Le fichier VCF que nous voulons annoter

Un paramètre supplémentaire à ajouter à notre commande est Xmx8G qui correspond au paramètre Java pour définir la mémoire disponible. La commande finale ressemblera à ceci :

Java -Xmx8g -jar snpEff/snpEff.jar NC_045512.2 (seq).vcf > (seq.final).vcf

7 Création du génome consensus

Pour reconstruire la séquence du génome de l'échantillon à partir du fichier VCF, nous pouvons utiliser bcftools qui nous génère un fichier Fasta:

tabix (seq. final). vcf.gz

bcftools consensus -f genome/reference.fasta (seq. final). vcf.gz > (sequence). fa

L'analyse génomique de la séquence issue du séquenceur Ion Torrent se base sur l'analyse d'un fichier Fasta et non pas Fastq. Pour cela quelques modifications ont été apporté à notre workflow (Annexe Figure 2) :

- L'outil d'alignement et d'indexation du génome de référence utilisé est minimap2 qui permet d'aligner des génomes complets.
- L'étape de marquage des duplicates et de création du génome consensus ont été supprimer.

III Analyse phylogénétique :

1 Collecte des données :

La structure GISAID (Global Initiative on Sharing Avian Influenza Data), qui a été élaborée pour collecter les données génomiques de la grippe, est actuellement engagée dans la collecte de données sur SARS-CoV-2. Cette structure (www.gisaid.org) permet une protection de la propriété intellectuelle qui est reconnue par la communauté internationale.

La collecte des données est mise en œuvre grâce à la base de données en ligne *Global Initiative on Sharing Avian Influenza Data* (GISAID), on se basant sur 3 critères : hôte (homme), la séquence complète et le variant d'omicron. Notre data set finale compte 864 génomes issus de patients Covid-19 diagnostiqués au monde entier.

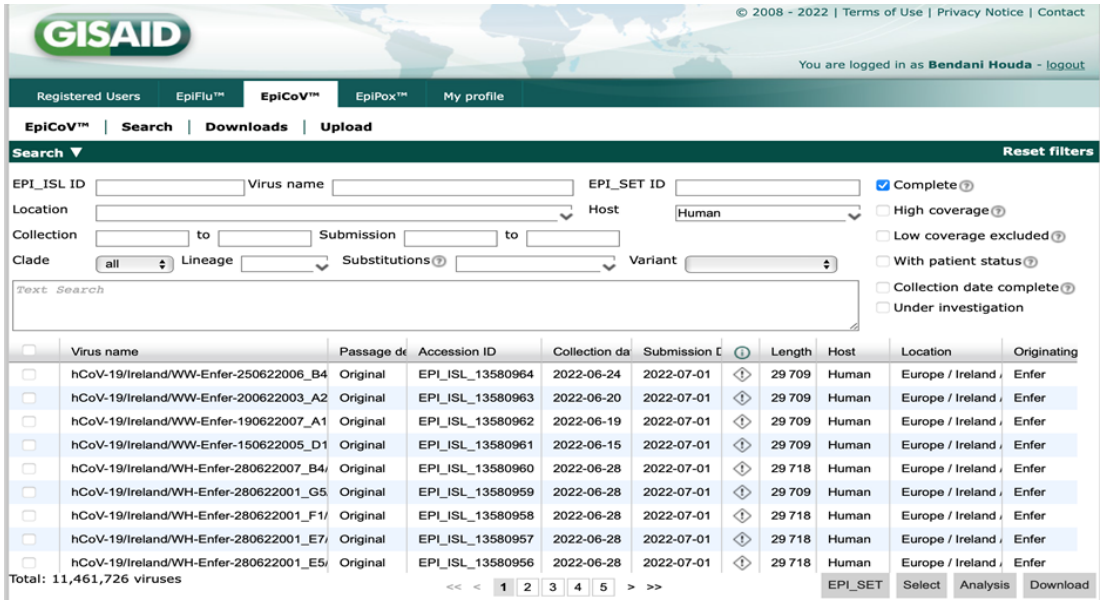


Figure 20: Interface de la plateforme Gisaid.

2 Contrôle de qualité :

Le contrôle de qualité a été effectué en utilisant Nextclade Web qui est disponible en ligne via l'adresse clades.nextstrain.org (66). Cette application accepte les données de séquence au format FASTA, qui assure l'alignement, la détection des mutations, l'attribution de clades, le placement phylogénétique, le contrôle de qualité et affiche les résultats sous forme de tableau ainsi que sous la forme de l'arbre phylogénétique. Les résultats peuvent également être téléchargés sous forme de fichiers, pour une analyse plus approfondie. Nextclade est conçu pour

des réponses rapides. L'ensemble de l'analyse, selon le nombre de séquences à traiter, prend de quelques secondes à quelques minutes.

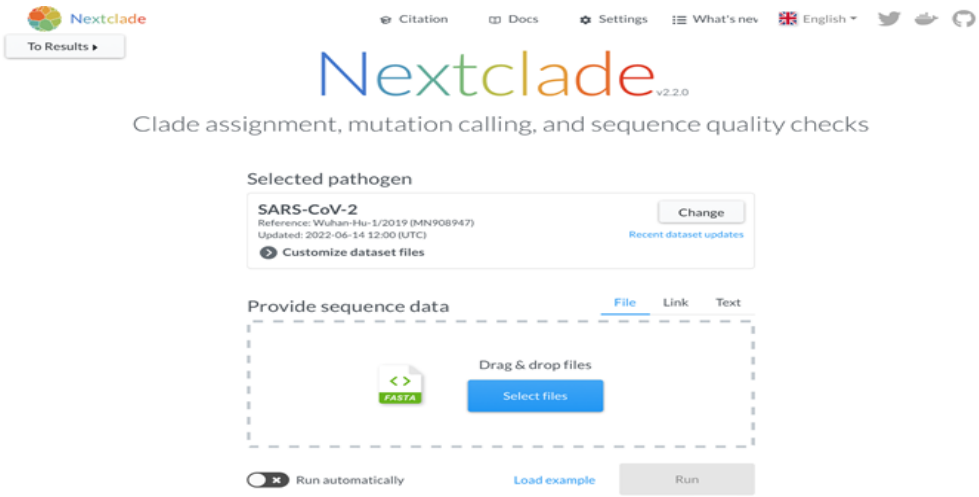


Figure 21: Interface de l'application Nextclade.

Le tableau des résultats s'affiche en outre pour chaque séquence :

- « M » : nombre de nucléotides ambigus et qui ne sont pas N
- « N » : nombre de nucléotides manquants indiqué par N
- « P » : nombre de nucléotides insérés par rapport à la séquence de référence
- « C » : amas de mutation affiché lors de la détection de plusieurs mutations
- « F » : nombre de décalages de frame inhabituels (le nombre total, y compris les décalages de frame courants)
- "S" : nombre de codons d'arrêt prématurés rares (le nombre total, y compris les arrêts prématurés courants)

Pour chaque séquence, chaque règle de QC produit un score de qualité. Ces scores de qualité individuels sont représentés dans le tableau suivant :

Tableau 2: Score de qualité utilisé par NextClade

Couleur affichée	Signification	Score
Vert	Bonne qualité	0 à 29
Jaune	Moyenne qualité	30 à 99
Rouge	Mauvais qualité	Supérieur ou égale à 100

3 Alignement de séquences multiples (MSA)

L'alignement de séquences multiples (MSA) est une étape importante dans les analyses comparatives de séquences biologiques (67). Plusieurs outils d'alignement de séquences multiples sont disponibles : Clustal Omega, MView, MAFFT, T-Coffe, …

Dans notre étude, nous avons choisi l'outil MAFFT qui dispose d'un service en ligne pour le calcul des MSA. Ce programme offre diverses options pour calculer de grands MSA composés de milliers de séquences, ajouter des séquences non alignées dans un alignement existant, l'ajustement de la direction dans l'alignement des nucléotides, l'alignement contraint et le traitement parallèle. Et qui contribue également à des fonctions supplémentaires telles que la sélection interactive de séquences et l'inférence phylogénétique afin de réaliser le prétraitement et le post-traitement de la MSA.

Toutes les options peuvent être sélectionnées dans une page pour un alignement large sur le serveur MAFFT (http://mafft.cbrc.jp/alignment/server/large.html). On a opté pour travailler avec les options par défaut sauf au niveau des paramètres de la matrice de notation des séquences nucléotidiques, ou on a choisi la valeur '1PAM / κ=2' qui est recommandée pour l'alignement de séquences d'ADN étroitement liées.

4 Inférence de l'arbre phylogénétique

PhyML est un logiciel de phylogénie basé sur le principe du maximum de vraisemblance (Maximum Likelihood ML), qui assure la comparaison entre les séquences génétiques afin d'établir leurs liens de parentés évolutives représentés sous forme des arbres phylogénétiques et pour expliquer l'évolution qui a permis l'apparition du SARS-CoV-2. Partout au monde des équipes s'en emparent pour remonter le fil de l'épidémie. Ce logiciel fait partie de l'ensemble d'outils disponibles afin de mieux comprendre et combattre le coronavirus SARS-CoV-2, responsable de l'épidémie COVID-19 (68). Pour utiliser PhyML, il suffit de mettre les séquences sur le site (http://www.atgc-montpellier.fr/sms/).

L'une des premières étapes de l'analyse phylogénétique est la sélection de modèles à l'aide de critères basés sur la vraisemblance. Il faut choisir à la fois une matrice de substitution et un modèle pour les taux entre les sites (RAS). Une méthode simple consiste à tester toutes les combinaisons et à sélectionner la meilleure. Un nouvel outil logiciel pour accomplir cette tâche est le SMS, qui signifie « Smart Model Selection ». Cet outil est très simple à utiliser et est

intégré dans l'environnement PhyML comme il peut également être utilisé en tant qu'application autonome en téléchargement (http://www.atgc-montpellier.fr/sms/)(69).

Ensuite on a accédé au FigTree à travers le site web (http://tree.bio.ed.ac.uk/software/figtree/), afin de visualiser notre arbre phylogénétique sous forme de graphe. Ce programme est destiné pour afficher des fichiers résumés et annotés. Il dispose d'une interface graphique qui permet aux utilisateurs de modifier divers composants de l'arbre tels que les positions d'enracinement, les étiquettes de pointe, les étiquettes de nœud et les axes d'échelle. Les figures d'arbres qui peuvent être exportées au format PDF afin d’être publiées ou modifiées ultérieurement dans un autre programme graphique.

Résultats

IV Analyse génomique

1 Alignement : fichier SAM

Le format de notre fichier SAM est Séquence Alignement/Map, ainsi qu'il convient aux reads courts et longs (Illumina, AB/Solid et Roche/454), et qui peut être utilisé comme fichiers de sortie par le projet 1000 génomes. Il est considéré comme un fichier texte tabulé (SAM), et contient deux sections : Entête (optionnel) et alignement. Concernant l'entête qu'on le trouve au début du fichier et qui commence par @, on trouve :

@HG : qui nous renseigne sur la version du format.

@SQ : identifie la liste des séquences de référence (une ligne par séquence de référence utilisée).

@PG : nom du programme.

@RG : est le groupe de lecture.

Chaque ligne d'alignement comporte 11 champs obligatoires pour les informations essentielles sur l'alignement, telles que la position de la carte, et un nombre variable de champs facultatifs pour des informations flexibles ou spécifiques à l'aligneur (tableau 2).

Tableau 3: description des 11 champs du fichier SAM.

Position	Champ	Description
1	QNAME	Nom du modèle de requête
2	FLAG	Drapeau au niveau du bit (appariement, brin, brin compagnon ...)
3	RNAME	Nom de la séquence de référence
4	POS	Position la plus à gauche de base 1 de l'alignement écrêté
5	MAPQ	Qualité de la cartographie
6	CIGAR	Ficelle de cigare allongée
7	MRNM	Nom de référence du compagnon (‘=’ si identique à RNAME)
8	MPOS	Position de contrainte la plus à gauche de base 1
9	ISIZE	Taille d'insertion déduite
10	SEQ	Séquence de requête sur le même brin que la référence
11	QUAL	Qualité de la requête

2 Appel des variations : VCF

Le fichier VCF a une section d'en-tête de 33 lignes. La première ligne du fichier indique que le format de ce fichier est la version 4.2 de VCF. Une autre ligne importante dans l'en-tête est la ligne qui spécifie que le génome de référence est reference. Fasta (Figure 24 ligne 5).

```
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.10.2+htslib-1.10.2-3
##bcftoolsCommand=mpileup -f reference.fasta -o inter.bcf sample.bam
##reference=file://reference.fasta
##contig=<ID=NC_045512.2,length=29903>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of raw reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of raw reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts
in RNA-seq data (bigger is better)",Version="3">
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SGB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=ICB,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
##INFO=<ID=HOB,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.10.2+htslib-1.10.2-3
##bcftools_callCommand=call --ploidy 1 -mv -o last.vcf inter.bcf; Date=Mon Jul  4 19:43:46 2022
##SnpEffVersion="5.1 (build 2022-01-21 06:23), by Pablo Cingolani"
##SnpEffCmd="SnpEff NC_045512.2 last.vcf "
##INFO=<ID=ANN,Number=.,Type=String,Description="Functional annotations: 'Allele | Annotation | Annotation_Impact | Gene_Name | Gene_ID
 | Feature_Type | Feature_ID | Transcript_BioType | Rank | HGVS.c | HGVS.p | cDNA.pos / cDNA.length | CDS.pos / CDS.length |
 AA.pos / AA.length | Distance | ERRORS / WARNINGS / INFO' ">
##INFO=<ID=LOF,Number=.,Type=String,Description="Predicted loss of function effects for this variant. Format: 'Gene_Name | Gene_ID
| Number_of_transcripts_in_gene | Percent_of_transcripts_affected'">
```

Figure 22: Capture de l'entête du fichier VCF.

L'en-tête de mon fichier VCF suivi de 12 lignes qui représentent chaque variant. Chacune contient :

- CHROM : qui nous renseigne sur le numéro de chromosome de la mutation.
- POS : indique la position de la mutation sur le chromosome
- ID : désigne le RSID de la mutation, c'est-à-dire un nom qui lui est donné pour le référencer. Dans mon fichier VCF, aucun RSID n'est donné et l'ID est affiché sous la forme d'un point sur chaque ligne. Ce n'est pas un problème, car la plupart des correspondances ADN sont effectuées par position, et non par les RSID qui peuvent changer de position entre les versions.
- REF : montre la valeur de cette position sur ce chromosome dans le génome de référence et est l'une des valeurs A, C, G et T.
- ALT : révèle sur les valeurs alternatives. Habituellement, il s'agit d'une valeur, l'une d'A, C, G et T et qui est différente de la valeur REF.
- QUAL : est un nombre estimant la qualité de la lecture qui a été effectuée dans le fichier. Lorsque le nombre est plus élevé c'est-à-dire qu'il est d'une meilleure qualité.
- FILTER : est une évaluation pour savoir si cette valeur est fiable.

```
#CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  sample.bam
NC_045512.2 241 .   C   T   225 .   DP=245;VDB=3.82373e-06;SGB=-0.693147;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,154,64;MQ=60
    GT:PL   1:255,0
```

Figure 23: Entête et première ligne du tableau de variantes du fichier VCF non annoté.

Le premier est un polymorphisme SNP de C/T → {C, T} → C est l'allèle de référence

NC_045512.2	241	.	C	T	225	.

Figure 24: Première ligne du tableau des variantes.

Le deuxième est un polymorphisme SNP de G/A → {G, A} → G est l'allèle de référence

NC_045512.2	1589	.	G	A	228	.

Figure 25: Deuxième ligne du tableau des variantes.

Le troisième est un polymorphisme SNP de C/T → {C, T} → C est l'allèle de référence

Figure 26: Troisième ligne du tableau des variantes.

Une autre colonne, INFO (Informations supplémentaires), contient des informations détaillées sur la lecture. Les champs INFO sont codés sous la forme d'une série de touches courtes séparées par des points-virgules avec des valeurs facultatives au format : <clé>=<données> [, données] :

- AA : Allèle ancestral.
- AC : nombre d'allèles dans les génotypes, pour chaque allèle ALT, dans le même ordre que celui indiqué.
- AF : Fréquence de l'allèle pour chaque allèle ALT dans le même ordre que celui indiqué.
- AN : nombre total d'allèles au sein des génotypes.
- DP : Profondeur combinée à travers les échantillons, par ex. DP=245.
- ANN : Annotations fonctionnelles. Ce sous champ ANN (Figure 29) dans le champ INFO du VCF inclus :
    - Allèle : A, T, C, G.
    - Annotation : missense variant, synonyme, splice site …
    - Annotation Impact : une estimation de l'impact : MODIFIER, FAIBLE, MODÉRÉ, ÉLEVÉ.
    - Gene Name : Nom commun du gène (HGNC), dans le cas de notre fichier VCF on trouve par exemple : S.
    - Gene ID : identifiant d'ensemble (exemple de notre fichier VCF :NC_045512.2)
    - Feature Type : transcrit, motif, miRNA …
    - Feature ID
    - Transcript Biotype : protéine codée, non codée …
    - Rang : Rang intron ou exon / nombre total d'introns et d'exons.
    - PL.c : c.1841A>G
    - PL. p : p. Asp614Gly

- cDNA.pos /cDNA.longueur : 1841/3822
- CDS.pos / CDS.longueur : 1841/3822
- AA.pos / AA. Longueur : 614/1273
- ERREURS / AVERTISSEMENTS / INFOS

```
ANN=G|missense_variant|MODERATE|S|GU280_gp02|transcript|GU280_gp02|protein_coding|1/1|c.1841A>G|p.Asp614Gly|1841/3822|1841/3822|614/1273||
```

Figure 27: Champs ANN du fichier VCF.

3 Identification des mutations

3.1 Séquence d'Illumina MiSeq (Fastq) :

Après avoir identifié la lignée de notre séquence d'Illumina qui est la lignée B.1.1 grâce à l'application pangolin, on a suivi l'analyse génomique de notre séquence fastq issue du séquenceur Illumina Miseq, comparé à la séquence de référence du SARS-cov-2. Toutes les mutations trouvées sont des SNPs (single nucleotide polymorphism) et aucune délétion ni insertion n'ont été identifié dans notre séquence.

Nous avons abouti à 12 mutations réparties dont 1 non codante. Pour les 11 mutations codantes, 3 sont synonymes et 8 mutations non Synonymes. Dont on trouve 4 mutations au niveau du gène ORF1-ab (G442S, T1055I, P4715L et M5060L), une seule au niveau de la protéine S(D614G) et 3 mutations au niveau de la protéine N (S202T, R203K et G204R) indiquées au niveau du tableau ci-dessous.

Quant aux 3 mutations codante synonymes avec un faible impact, deux mutations au niveau du gène ORF1-ab situées aux positions 3037 et 11974 respectivement et une mutation au niveau du gène N à la position 28882.

.

Tableau 4: Mutations non synonymes codantes extraites de la séquence d'Alpha (fastq).

pos	ref	alt	type_mutations	gene_name	mutation-nt	mutation-aa	peptide
1589	G	A	missense_variant	ORF1ab	1324G>A	G442S	NSP2
3429	C	T	missense_variant	ORF1ab	3164C>T	T1055I	NSP3
14408	C	T	missense_variant	ORF1ab	14144C>T	P4715L	NSP12
15442	A	T	missense_variant	ORF1ab	15178A>T	M5060L	NSP12
23403	A	G	missense_variant	S	1841A>G	D614G	
28878	G	C	missense_variant	N	605G>C	S202T	
28881	G	A	missense_variant	N	608G>A	R203K	
28883	G	C	missense_variant	N	610G>C	G204R	

3.2 **Séquence d'Ion torrent (Fasta) :**

D'après l'analyse génomique de notre deuxième séquence FASTA qui est de la lignée BA.1 dévoilée par le site suivant :https://pangolin.cog-uk.io/, notre séquence FASTA comporte 60 mutations.

A partir de ces derniers, 6 mutations sont des délétions représentées dans le tableau 4, dont une délétion a un impact élevé qui touche l'expression de la protéine M tandis que les autres délétions codent 2 protéines ORF1ab, 2 protéines S et une seule protéine N.

Les mutations dépourvues de délétions (SNP) sont au nombre de 54, dont deux mutations qui ne sont pas codantes. Pour les 52 autres mutations codantes, 9 sont synonymes avec un impact faible. Alors que le reste des mutations codantes non synonymes sont au nombre de 43 qui codent pour 7 protéines ORF1-ab, 29 protéines S ,3 protéines M, 3proteines N et une seule protéine E. (Annexe Tableau 2)

Tableau 5: Délétion de la séquence Omicron (FASTA).

pos	ref	alt	type_mutations	anno_impact	gene	mutation-aa
6512	AGTT	A	disruptive_inframe	MODERATE	ORF1ab	Ser2083_Leu2084delinsIle
11282	AGTTT GTCTGGTTT	AGTTT	disruptive_inframe	MODERATE	ORF1ab	Leu3674_Gly3676del
21764	ATACATGT	AT	disruptive_inframe	MODERATE	S	His69_Val70del
21986	GGTGT TTATT	G	disruptive_inframe	MODERATE	S	Gly142_Tyr145delinsAsp
26625	CTACAA TTTGCCTAT	CT	frameshift_variant	HIGH	M	Q36None
28361	GGAGA ACGCAG	GG	disruptive_inframe	MODERATE	N	Glu31_Ser33del

V Analyse phylogénétique

1 Contrôle de qualité

Afin de réaliser un alignement de 864 séquences au format fasta avec une séquence de référence on a accédé vers l'application web Nextclade afin de nous identifier plusieurs mesures de contrôle de la qualité tels que l'attribution des séquences à un clade ou une variante dans le but de détecter les changements et les mutations des protéines génomiques par rapport à la séquence de référence ce qui aboutit d'établir une analyse plus facile et rapide. Et même d'afficher les résultats sous forme de tableau, d'arbre phylogénétique dans cette notre étude on a choisi de télécharger les résultats sous forme de fichiers CSV afin d'assurer un examen et une analyse plus approfondis.

Vous pouvez obtenir un aperçu rapide de l'écran des résultats dans la figure 31 ci-dessous :

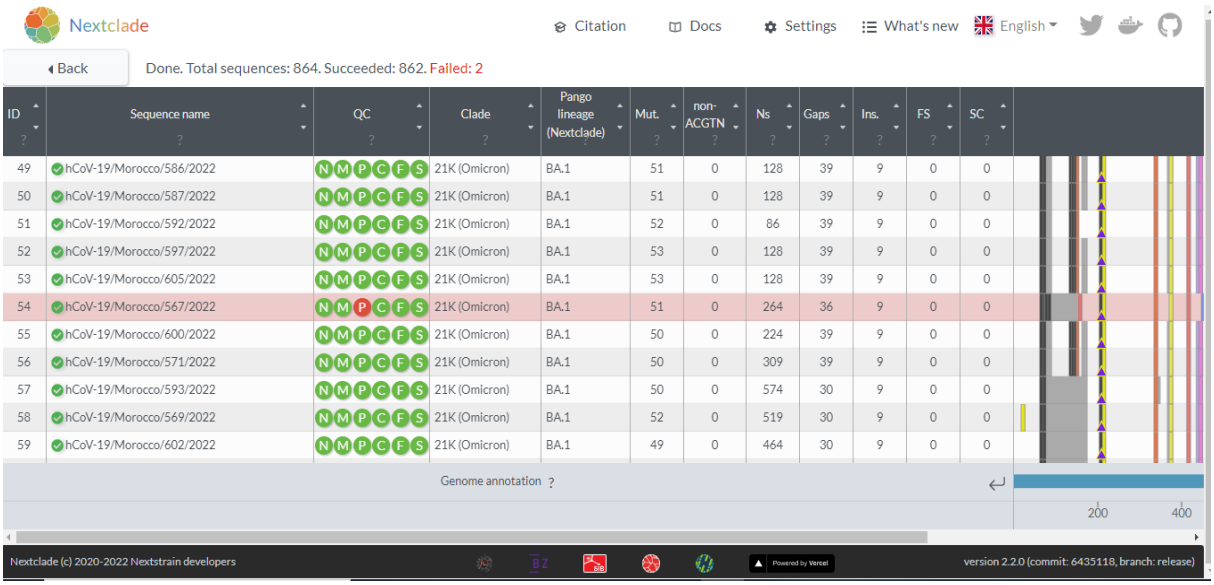


Figure 28: résultats du contrôle qualité des séquences de l'analyse phylogénétique par l'application Nextclade.

Après l'exécution de l'analyse par Nextclade, il nous a mis en œuvre une variété de mesures de contrôle qualité pour détecter rapidement les problèmes dans notre assemblage et évaluer les séquences qui posent problème en triant le tableau des résultats de mauvais à bon, Les mauvaises séquences sont colorées en rouge, les médiocres en orange et les bonnes en vert comme si indiquées en pourcentage des 3 situations dans le graphe de la figure 32. Mais après le filtrage, notre data set comportait juste les 281 séquences de bonne qualité représentant 37%, et distribuées en divers pays dont on trouve 52 séquences au Maroc, 89 séquences issues d'USA etc. Comme si montré dans la figure 33.

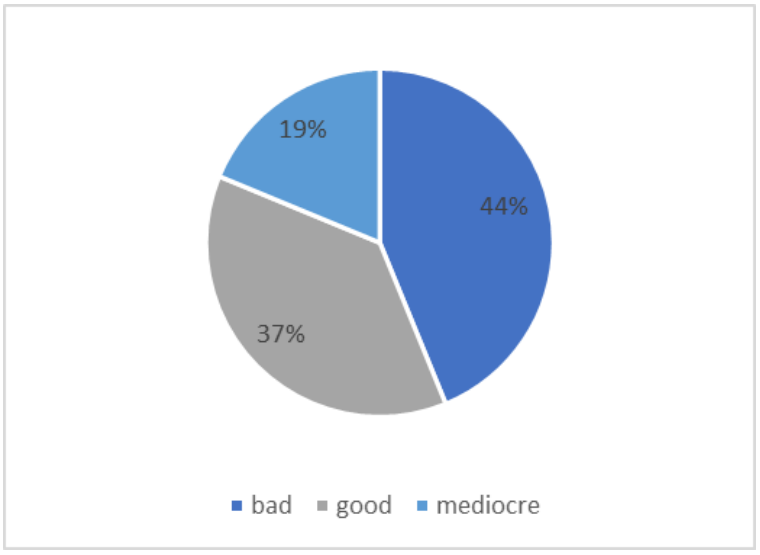


Figure 29:Diagramme illustrant les pourcentages de la qualité des séquences restantes après le filtrage.

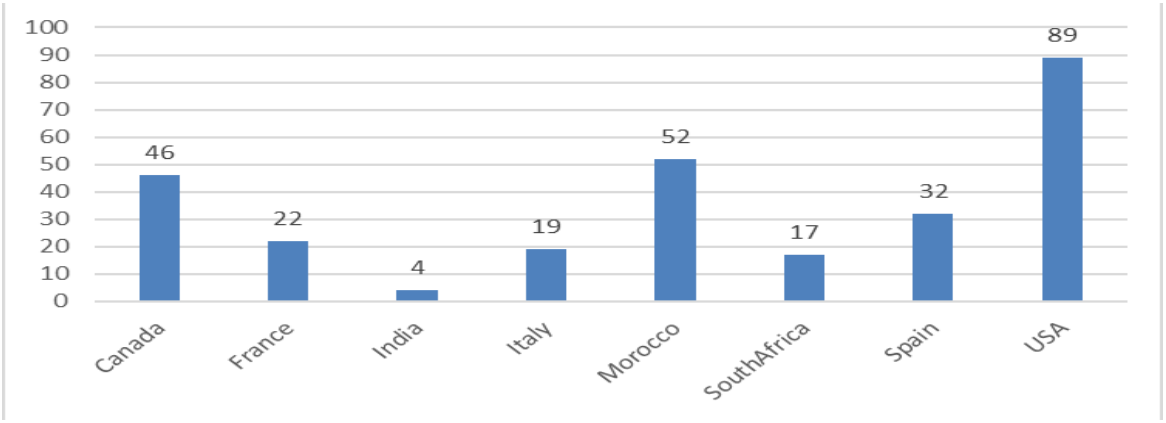


Figure 30:Représentation graphiques de la distribution des séquences omicron au sein de notre data set

2 Phylogénie

On a présenté les 281 séquences restantes au programme d'alignement multiple MAFFT qui a supprimé le sur-alignement de 34 séquences après avoir défini le seuil de suppression des lettres ambiguës en plus de 5%. Pour les 243 séquences restantes, elles ont été utilisées en plus de la référence afin de nous servir à créer l'arbre phylogénétique en employant le logiciel "Smart Model Selection" (SMS) implémenté dans l'environnement PhyML qui nous a sélectionné le model TN93+R comme étant le modèle adapté à nos données comme indiqué dans la figure 33 ci-dessous.

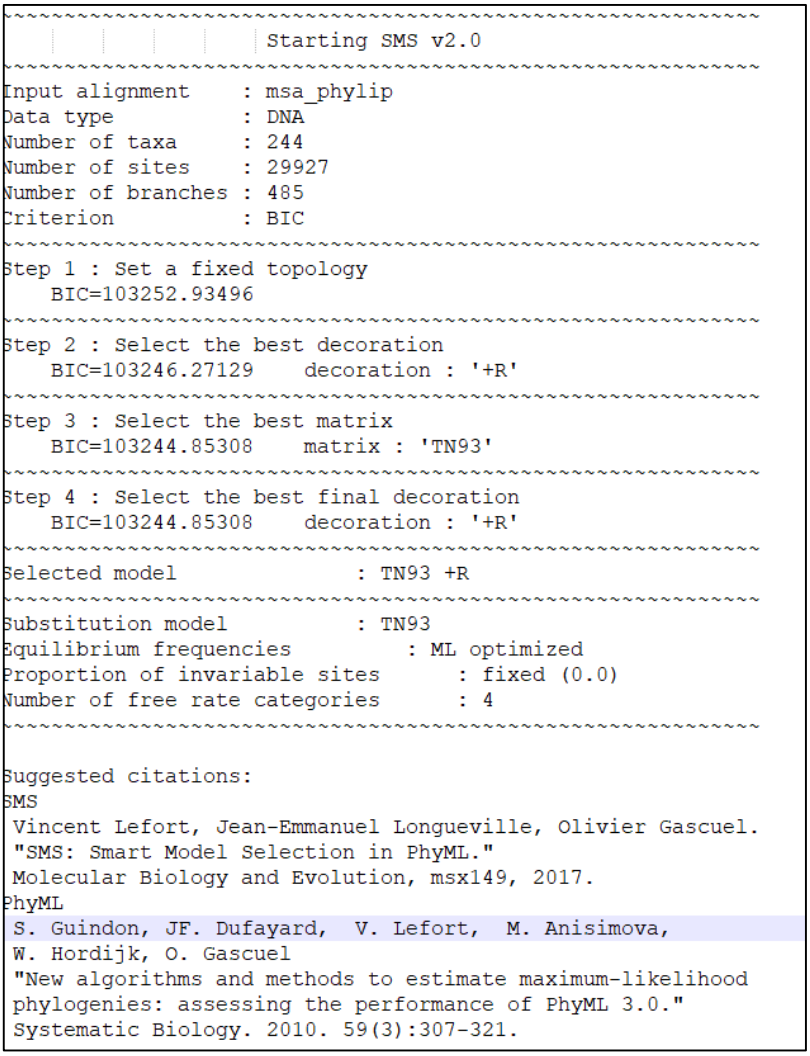


Figure 31: output de la sélection du model réalisé par SMS.

L'arbre phylogénétique complète est représenté dans la figure 35. Nous avons sélectionné en bleu quelques clades qui comprennent des séquences Marocaines.

Dans le cas de cette analyse, il a été constaté que la majorité des séquences sont regroupés au sein des clades hétérogènes.

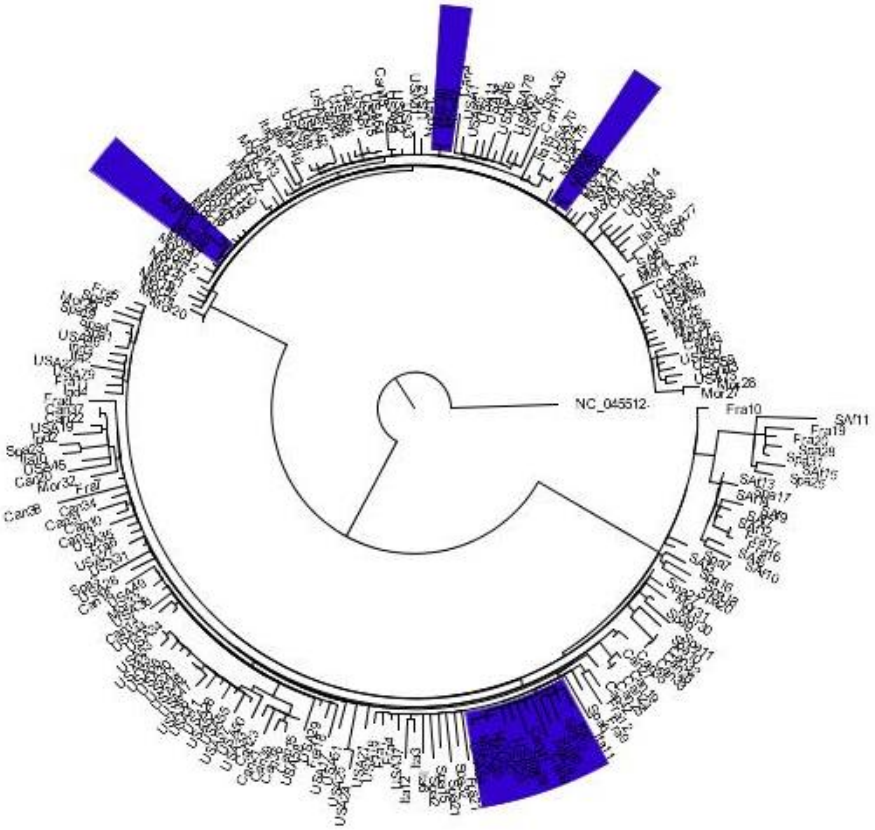


Figure 32: Arbre phylogénétique des 243 séquences.

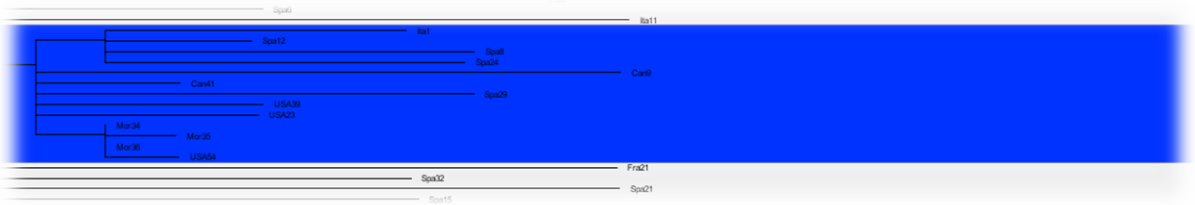


Figure 33: Partie de l'arbre phylogénétique représentant le clade (A).

La figure ci-dessus représente le plus gros clade qui contient des séquences de personnes infectées provenant de différents pays tels que le Maroc, Espagne, Canada, USA et l'Italie. On constate aussi que les séquences du Maroc renferment un petit nombre de mutations par rapport aux autres séquences.

Figure 34: Partie de l'arbre phylogénétique représentant le clade (B).

On note toutefois aussi une structure locale du variant BA.1 au Maroc dans le clade B représenté dans la figure 36. Cette partie du clade (en bleu) comprends 4 séquences dont la longueur de la branche diffère. Ceci veut dire qu'il y a une différence dont le nombre de mutations. Par exemple au sein du groupe, la séquence Maroc 24 comporte plus de mutations par rapport à la séquence Maroc 13.

Figure 35: Partie de l'arbre phylogénétique représentant le clade (C).

Ce troisième clade (Figure 37) renferme 3 branches correspondant à des séquences issues de la France, du Maroc et de l'Italie avec un nombre de mutations plus ou moins proche entre le groupe de l'Italie 18 et la France 22.

DISCUSSION

L'Omicron a été repéré pour la première fois au Botswana et en Afrique du Sud et s'est étalé rapidement à travers l'Afrique du Sud et partout dans le monde. Au 16 décembre 2021, il était présent sur 5 continents dont l'Europe, l'Amérique du Nord, l'Amérique du Sud, l'Asie, l'Afrique et l'Océanie, et au moins dans 77 pays dont le Maroc, les États-Unis, le Royaume-Uni, la France, l'Allemagne, l'Australie, le Japon et La Chine a tous signalé des cas d'Omicron. Cela a sonné l'alarme pour de nombreux pays et régions du monde, provoquant la peur et l'incertitude dans la lutte mondiale contre la pandémie de COVID-19. L'émergence d'Omicron est un autre coup désagréable pour l'économie mondiale qui est déjà en difficulté, ce qui ajoutera certainement plus d'incertitudes pour nous tous à travers le monde (71).

La troisième vague du covid-19 (Omicron) a connu une propagation massive à l'échelle nationale avec un pourcentage d'infection ait atteint 98% du total des cas enregistrés de la semaine 13 au 19 décembre.

Le variant Omicron dispose maintenant de cinq sous-lignages différents, qui sont apparues dans les séquences chronologiques suivantes : BA1, BA2, BA3 puis BA4 et BA5", présents en Afrique du Sud et dans plusieurs autres pays. BA1, BA2 et BA3 ont plusieurs mutations communes. BA1 et BA2 ont, en plus, leurs mutations spécifiques. Et BA3 est entre les deux : il prend des mutations de BA1 et BA2, mais il n'a aucune mutation qui lui soit propre. Concernant la vague Omicron BA1 est arrivé, on a pu rapidement distinguer les variantes Omicron et Delta, parce qu'une des mutations présentes dans BA1 est une délétion au niveau du gène Spike.

On rappelle que les mutations non synonymes modifient les séquences protéiques et sont généralement soumises à la sélection naturelle. C'est tout à fait pareil des mutations non-sens qui introduisent des codons stop prématurés au sein des séquences codantes, il y a généralement une insertion ou une suppression d'un seul nucléotide dans la séquence pendant la transcription lorsque l'ARN messager copie l'ADN. Ce qui provoque une mutation de décalage de cadre qui perturbe tout le cadre de lecture de la séquence d'acides aminés et mélange les codons. Cela affecte généralement les acides aminés qui sont codés et modifie la protéine résultante qui est exprimée (72).

Pourtant Les mutations synonymes sont considérées fonctionnellement silencieuses et évolutivement neutres. Autrement dit la majorité des acides aminés ont plusieurs codons d'ARN qui se traduisent en acide aminé particulier ce qui explique le cas de la mutation
Synonyme par exemple si le troisième nucléotide du codon est celui qui porte la mutation il en résulte un codage pour le même acide aminé c'est à dire le codon muté à la même signification que le codon d'origine et si l'acide aminé ne change pas la protéine également ne se modifie plus. Cela signifie que les mutations synonymes n'ont aucun effet et par conséquent ne sont pas remarquées (73).

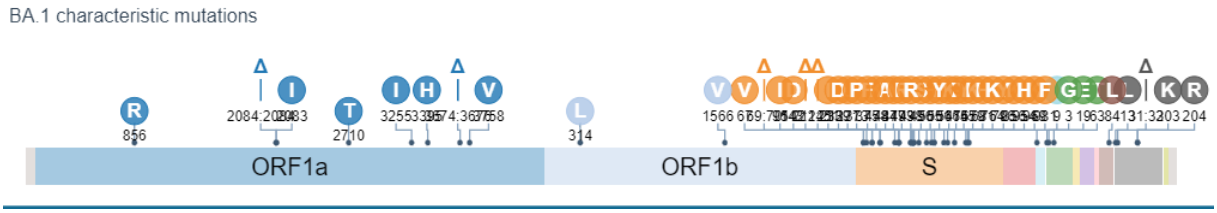


Figure 36: Mutations de référence du variante BA.1 (Outbreak.info).

Comparé à la séquence de référence du SARS-CoV-2, le variant Omicron comporte 49 substitutions d'acides aminés et 6 délétions. La protéine S seule rassemble 30 de ces substitutions avec 3 délétions.

Tableau 6: Délétion de référence du variant BA.1 (Outbreak.info).

Gene	Mutation
ORF1a	del2084/2084
ORF1a	del3674/3676
S	del69/70
S	del143/145
S	del212/212
N	del31/33

La moitié de ces mutations concerne le receptor binding domain (RBD) qui interagit avec le récepteur ACE 2. L'analyse des précédentes variantes suggère que les mutations du RBD modifient significativement le comportement du coronavirus (74).
Plus précisément, neuf acides aminés semblent accroître les liens entre la protéine S et le récepteur ACE2, et donc sa capacité à infecter les cellules. C'est le cas des 4 mutation, qui sont identifié dans la variante Omicron, suivantes : K417N, Q493R, N501Y, et Y505H. La mutation N501Y, observée chez d'autres variantes, et qui augmente l'affinité de la protéine (S) avec le

récepteur ACE2 (71,68). Les autres mutations situées en amont ou en aval du RBD pourraient aussi participer au succès épidémiologique du variant Omicron. Des modifications à proximité du site de coupure des deux sous-unités de la protéine S pourraient consolider sa résistance aux anticorps neutralisants. La mutation N679K, une des trois identifiées, n'a jamais été observée ailleurs.

Au-delà de la protéine S, deux mutations dans la nucléocapside (N), la coque protectrice du génome du virus, sont aussi présentes. Elles sont connues pour amplifier la réplication de l'ARN viral ce qui peut provoquer le déclenchement d'un nouveau pic épidémique mondial.

Ainsi que l'émergence du sous variant BA.2. Il faut prendre en compte que la vague d'Omicron a commencé au début de l'année 2022. Durant cette période, c'était le sous-variant BA.1 qui était prédominant. Mais, par la suite, au fur et à mesure que la vague d'omicron se dissipait, c'était le BA.2 qui s'est greffé mais il y a également le BA.5, qui est détecté au Maroc actuellement. Il va remplacer le BA.2 par la suite pour la simple raison qu'il est encore plus transmissible que lui.

Conclusion

Cette étude avait pour objectif d'analyser la variante omicron au Maroc. Dans ce sens, on s'est focalisé sur l'analyse de là sous variante BA.1 (puisqu'elle était prédominant au moment de la rédaction du présent document). Nous avons d'abord effectué une analyse génomique d'une séquence marocaine appartenant à BA.1 afin d'identifier les différentes mutations (Délétion, Insertion et SNP). Ensuite, un nombre total de 864 séquences de BA.1 de différents pays (dont 52 du Maroc) ont été extraites et alignées sur la séquence de référence afin de générer un arbre phylogénétique.

Lors de cette analyse, nous avons constaté la présence de 60 mutations. Environ 75 % de ces mutations impactent des protéines structurelles (dont plus de 50 % concernent la protéine S**)**, les 25 % restant portant sur les protéines non structurelles (NSP, de régulation). De plus, 6 délétions sont également présentes, avec l'absence total d'insertion Plusieurs de ces mutations peuvent avoir des impacts sur la contagiosité, la virulence ou la résistance à l'immunité de ce variant. .

En termes de phylogénétique**,** on voit clairement qu'Omicron présente les séquences du Maroc en deux différents clades ; un clade séquentiel marocain exprimant l'évolution interne de la variante au Maroc avec la présence des clades hétérogènes démontrant que le profil de la variante BA.1 au Maroc est importé.

Finalement, nous sommes parvenus à révéler la diversité génétique et géographique de la variante BA.1, ainsi que la façon dont elle réduit l'efficacité des interventions disponibles.

Notre étude a identifié plusieurs clades et mutations qui peuvent être la base d'une étude ancestrale des origines spécifiques des mutations. Une étude approfondie regroupant un nombre important de séquences sera nécessaire pour démêler le profil de la variante BA.1 et son mode de transmission.

References

1. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res. juill 2020;24:91-8.

2. Helmy YA, Fawzy M, Elaswad 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. 24 avr 2020;9(4):1225.

3. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. Genomics. févr 2009;93(2):105-11.

4. Al-Qahtani AA. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): Emergence, history, basic and clinical aspects. Saudi J Biol Sci. oct 2020;27(10):2531-8.

5. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. mars 2019;17(3):181-92.

6. 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 [cité 7 oct 2022]. Disponible sur: https://www.ncbi.nlm.nih.gov/books/NBK554776/figure/article-52171.image.f3/

7. COVID-19 Data Explorer - Our World in Data [Internet]. [cité 1 nov 2022]. Disponible sur: https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&time=2020-03-01..latest&facet=none&pickerSort=asc&pickerMetric=location&Metric=Confirmed+cases&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false&country=USA~GBR~CAN~DEU~ITA~IND

8. Nouveau coronavirus (2019-nCoV) [Internet]. [cité 1 nov 2022]. Disponible sur: https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAjwh4ObBhAzEiwAHzZYU7k9FodxcfB6q-fIfxoBwTWkYy1DOM9hJUYMIYq3tCTi5cEg9O5cphoCZMEQAvD_BwE

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;42:1604061.

10. Shi Y, Wang G, Cai X peng, Deng J wen, Zheng L, Zhu H hong, et al. An overview of COVID-19. J Zhejiang Univ Sci B. mai 2020;21(5):343-60.

11. Pascarella G, Strumia A, Piliego C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: a comprehensive review. J Intern Med. 13 mai 2020;10.1111/joim.13091.

12. Comparing coronavirus to the flu and other respiratory illnesses [Internet]. [cité 28 juin 2022]. Disponible sur: https://www.news10.com/news/coronavirus/comparing-coronavirus-to-the-flu-and-other-respiratory-illnesses/

13. Belete TM. An Up-to-Date Overview of Therapeutic Agents for the Treatment of COVID-19 Disease . Clin Pharmacol Adv Appl. 14 déc 2020;12:203-12.

14. Basille D, Andrejak C. Infection à SARS-CoV-2 : connaissances au 15 avril 2021. Rev Mal Respir. juin 2021;38(6):616-25.

15. Ng TSB, Leblanc K, Yeung DF, Tsang TSM. Médicaments utilisés durant la COVID-19: Examen des données probantes récentes. Can Fam Physician Med Fam Can. mars 2021;67(3):e69-78.

16. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med. 1 oct 2019;200(7):e45-67.

17. Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial Version 7). Chin Med J (Engl). 5 mai 2020;133(9):1087-95.

18. Matusik É, Ayadi M, Picard N. Covid-19, prise en charge, pistes thérapeutiques et vaccinales. Actual Pharm. oct 2020;59(599):27-33.

19. Garnier M, Quesnel C, Constantin JM. Atteintes pulmonaires liées à la COVID-19. Presse Médicale Form. févr 2021;2(1):14-24.

20. Ng TSB, Leblanc K, Yeung DF, Tsang TSM. Médicaments utilisés durant la COVID-19. Can Fam Physician. mars 2021;67(3):e69-78.

21. Boopathi S, Poma AB, Kolandaivel P. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. J Biomol Struct Dyn. 30 avr 2020;1-10.

22. Huang Y, Yang C, Xu X feng, Xu W, Liu S wen. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. sept 2020;41(9):1141-9.

23. Lefeuvre C, Przyrowski É, Apaire-Marchais V. Aspects virologiques et diagnostic du coronavirus Sars-CoV-2. Actual Pharm. oct 2020;59(599):18-23.

24. Thomas S. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. Pathog Immun. 19 oct 2020;5(1):342-63.

25. Chang C ke, Chen CMM, Chiang M hui, Hsu Y lan, Huang T huang. Transient oligomerization of the SARS-CoV N protein--implication for virus ribonucleoprotein packaging. PloS One. 2013;8(5):e65045.

26. Histoire du COVID-19 – C5 : Infectiosité et réplication hors norme du SARS-Cov2 [Internet]. FranceSoir. [cité 7 oct 2022]. Disponible sur: https://www.francesoir.fr/societe-science-tech/histoire-du-covid-19-c5-infectiosite-et-replication-hors-norme-du-sars-cov2

27. Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, Dhamija P, et al. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. Cells. 6 avr 2021;10(4):821.

28. Michel CJ, Mayer C, Poch O, Thompson JD. Characterization of accessory genes in coronavirus genomes. Virol J. 27 août 2020;17:131.

29. Pyrć K. The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. :12.

30. 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 Otorhinolaryngol. août 2020;277(8):2251-61.

31. Astuti I, Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes Metab Syndr. 2020;14(4):407-12.

32. Kumar S, Nyodu R, Maurya VK, Saxena SK. Morphology, Genome Organization, Replication, and Pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Coronavirus Dis 2019 COVID-19. 30 avr 2020;23-31.

33. 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. 10 avr 2020;21(7):E2657.

34. 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. 15 juin 2020;35(Suppl 2):77.

35. Afzal A. Molecular diagnostic technologies for COVID-19: Limitations and challenges. J Adv Res. nov 2020;26:149-59.

36. Sule WF, Oluwayelu DO. Real-time RT-PCR for COVID-19 diagnosis: challenges and prospects. Pan Afr Med J. 2020;35(Suppl 2):121.

37. Zowawi HM, Alenazi TH, AlOmaim WS, Wazzan A, Alsufayan A, Hasanain RA, et al. Portable RT-PCR System: a Rapid and Scalable Diagnostic Tool for COVID-19 Testing. J Clin Microbiol. 20 avr 2021;59(5):e03004-20.

38. Munne K, Bhanothu V, Bhor V, Patel V, Mahale SD, Pande S. Detection of SARS-CoV-2 infection by RT-PCR test: factors influencing interpretation of results. VirusDisease. juin 2021;32(2):187-9.

39. Figure 9. The protocol template COVID-19 diagnostic testing through... [Internet]. ResearchGate. [cité 7 oct 2022]. Disponible sur: https://www.researchgate.net/figure/The-protocol-template-COVID-19-diagnostic-testing-through-real-time-RT-PCR-1_fig5_351514416

40. Adams SJ, Dennie C. L'imagerie thoracique chez les patients soupçonnés d'avoir la COVID-19. CMAJ Can Med Assoc J. 16 nov 2020;192(46):E1494.

41. Singh N, Fratesi J. TDM thoracique d'un des premiers cas canadiens de COVID-19 chez un homme de 28 ans. CMAJ Can Med Assoc J J Assoc Medicale Can. 19 oct 2020;192(42):E1286-7.

42. Lodé B, Jalaber C, Orcel T, Morcet-Delattre T, Crespin N, Voisin S, et al. Imagerie de la pneumonie COVID-19. J Imag Diagn Interv. sept 2020;3(4):249-58.

43. Desvaux É, Faucher JF. Covid-19 : aspects cliniques et principaux éléments de prise en charge. Rev Francoph Lab. nov 2020;2020(526):40-7.

44. Covid long. In: Wikipédia [Internet]. 2022 [cité 7 oct 2022]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=Covid_long&oldid=194994059

45. Hsieh WY, Lin CH, Lin TC, Lin CH, Chang HF, Tsai CH, et al. Development and Efficacy of Lateral Flow Point-of-Care Testing Devices for Rapid and Mass COVID-19 Diagnosis by the Detections of SARS-CoV-2 Antigen and Anti-SARS-CoV-2 Antibodies. Diagnostics. oct 2021;11(10):1760.

46. Van Caeseele P, Bailey D, Forgie SE, Dingle TC, Krajden M, pour l'Association pour la microbiologie médicale et l'infectiologie Canada, et al. Sérologie du SRAS-CoV-2 (COVID-19) : Répercussions sur la pratique clinique, la médecine de laboratoire et la santé publique. CMAJ Can Med Assoc J J Assoc Medicale Can. 7 déc 2020;192(49):E1776-82.

47. Bertholom C. Sars-CoV-2 : émergence, aspects virologiques et diagnostiques. Option/Bio. oct 2020;31(623):21-3.

48. Lamoril J, Ameziane N, Deybach JC, Bouizegarène P, Bogard M. Les techniques de séquençage de l'ADN : une révolution en marche. Première partie. Immuno-Anal Biol Spéc. oct 2008;23(5):260-79.

49. Lacoste C, Fabre A, Pécheux C, Lévy N, Krahn M, Malzac P, et al. Le séquençage d'ADN à haut débit en pratique clinique. Arch Pédiatrie. 1 avr 2017;24(4):373-83.

50. Crossley BM, Bai J, Glaser A, Maes R, Porter E, Killian ML, et al. Guidelines for Sanger sequencing and molecular assay monitoring. J Vet Diagn Investig Off Publ Am Assoc Vet Lab Diagn Inc. nov 2020;32(6):767-75.

51. Fig. 4 Steps of library preparation for Illumina sequencing (Zhou et... [Internet]. ResearchGate. [cité 8 oct 2022]. Disponible sur: https://www.researchgate.net/figure/Steps-of-library-preparation-for-Illumina-sequencing-Zhou-et-al-2010_fig3_338830137

52. Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, et al. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. BMC Genomics. 3 janv 2012;13:1.

53. 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. avr 2014;25(3-4):120-8.

54. 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.

55. SureshniFernando. Ion torrent sequencing [Internet]. 05:19:06 UTC [cité 8 oct 2022]. Disponible sur: https://pt.slideshare.net/SureshniFernando/ion-torrent-sequencing-226383768

56. Enrichissement ISP | biorigami [Internet]. [cité 8 oct 2022]. Disponible sur: http://www.biorigami.com/?tag=enrichissement-isp

57. 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. déc 2012;50(12):3831-7.

58. juin | 2013 | biorigami [Internet]. [cité 8 oct 2022]. Disponible sur: http://www.biorigami.com/?m=201306

59. Hsiao YP, Lu CT, Chang-Chien J, Chao WR, Yang JJ. Advances and Applications of Ion Torrent Personal Genome Machine in Cutaneous Squamous Cell Carcinoma Reveal Novel Gene Mutations. Mater Basel Switz. 14 juin 2016;9(6):E464.

60. MiniSeq - Séquenceur NGS de laboratoire by Illumina, Inc. | MedicalExpo [Internet]. [cité 1 nov 2022]. Disponible sur: https://www.medicalexpo.fr/prod/illumina-inc/product-83632-761729.html

61. Genomics - Ion Torrent DNA Sequencing Systems - Sunnybrook Research Institute [Internet]. [cité 1 nov 2022]. Disponible sur: https://sunnybrook.ca/research/content/?page=sri-core-genomics-equip-ion-torrent-dna-sequencing

62. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 15 juill 2009;25(14):1754-60.

63. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 1 nov 2011;27(21):2987-93.

64. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 1 févr 2021;10(2):giab008.

65. SnpEff and SnpSift [Internet]. [cité 14 juill 2022]. Disponible sur: http://pcingola.github.io/SnpEff/

66. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw. 30 nov 2021;6(67):3773.

67. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform | Nucleic Acids Research | Oxford Academic [Internet]. [cité 13 juill 2022]. Disponible sur: https://academic.oup.com/nar/article/30/14/3059/2904316

68. Stoner MCD, Dennis AM, Hughes JP, Eshleman SH, Sivay MV, Hudelson SE, et al. Characteristics Associated With Human Immunodeficiency Virus Transmission Networks Involving Adolescent Girls and Young Women in Human Immunodeficiency Virus Prevention Trials Network 068 Study. Sex Transm Dis. 1 mai 2019;46(5):e46-9.

69. Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 1 sept 2017;34(9):2422-4.

70. Amir IJ, Lebar Z, Yahyaoui G, Mahmoud M. Covid-19 : virologie, épidémiologie et diagnostic biologique. Option/Bio. août 2020;31(619):15.

71. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol. avr 2022;94(4):1641-9.

72. 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. mars 2022;100(3):217-21.

73. Araf Y, Akter F, Tang Y, Fatemi R, Parvez MdSA, Zheng C, et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. J Med Virol. mai 2022;94(5):1825-32.

74. 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. 7 janv 2022;10(1):1-11.

75. 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. 2 janv 2022;2021.12.20.21268048.

Annexe Tableaux

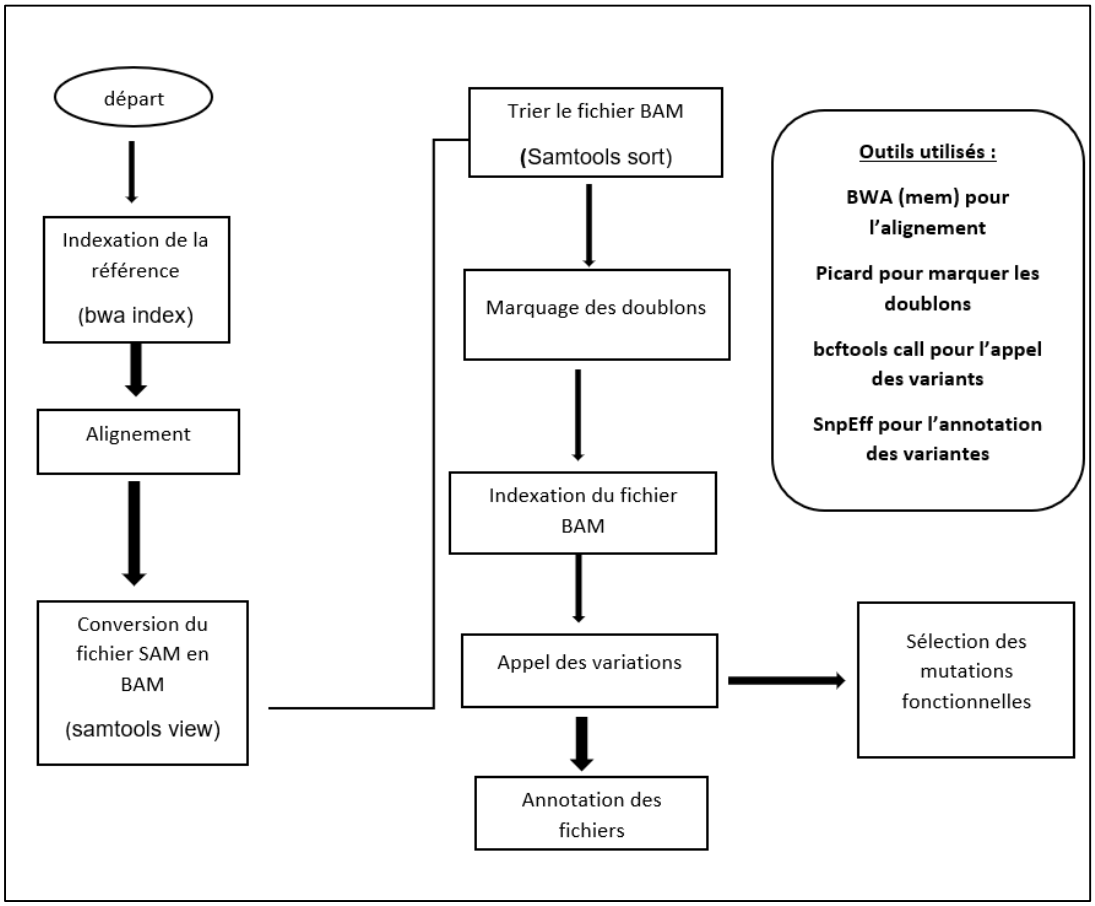
Annexe tableau 1: variantes identifiés depuis la séquence Alpha (FASTQ).

pos	ref	alt	type_mutations	impact	gene	mutation-nt	mutation-aa	peptide
241	C	T	intergenic_region	MODIFIER	CHR_START-ORF1ab	241C>T	Not found	Not found
1589	G	A	missense_variant	MODERATE	ORF1ab	1324G>A	G442S	NSP2
3037	C	T	synonymous_variant	LOW	ORF1ab	2772C>T	F924F	NSP3
3429	C	T	missense_variant	MODERATE	ORF1ab	3164C>T	T1055I	NSP3
11974	T	C	synonymous_variant	LOW	ORF1ab	11709T>C	D3903D	NSP7
14408	C	T	missense_variant	MODERATE	ORF1ab	14144C>T	P4715L	NSP12
15442	A	T	missense_variant	MODERATE	ORF1ab	15178A>T	M5060L	NSP12
23403	A	G	missense_variant	MODERATE	S	1841A>G	D614G	
28878	G	C	missense_variant	MODERATE	N	605G>C	S202T	
28881	G	A	missense_variant	MODERATE	N	608G>A	R203K	
28882	G	A	synonymous_variant	LOW	N	609G>A	R203R	
28883	G	C	missense_variant	MODERATE	N	610G>C	G204R	

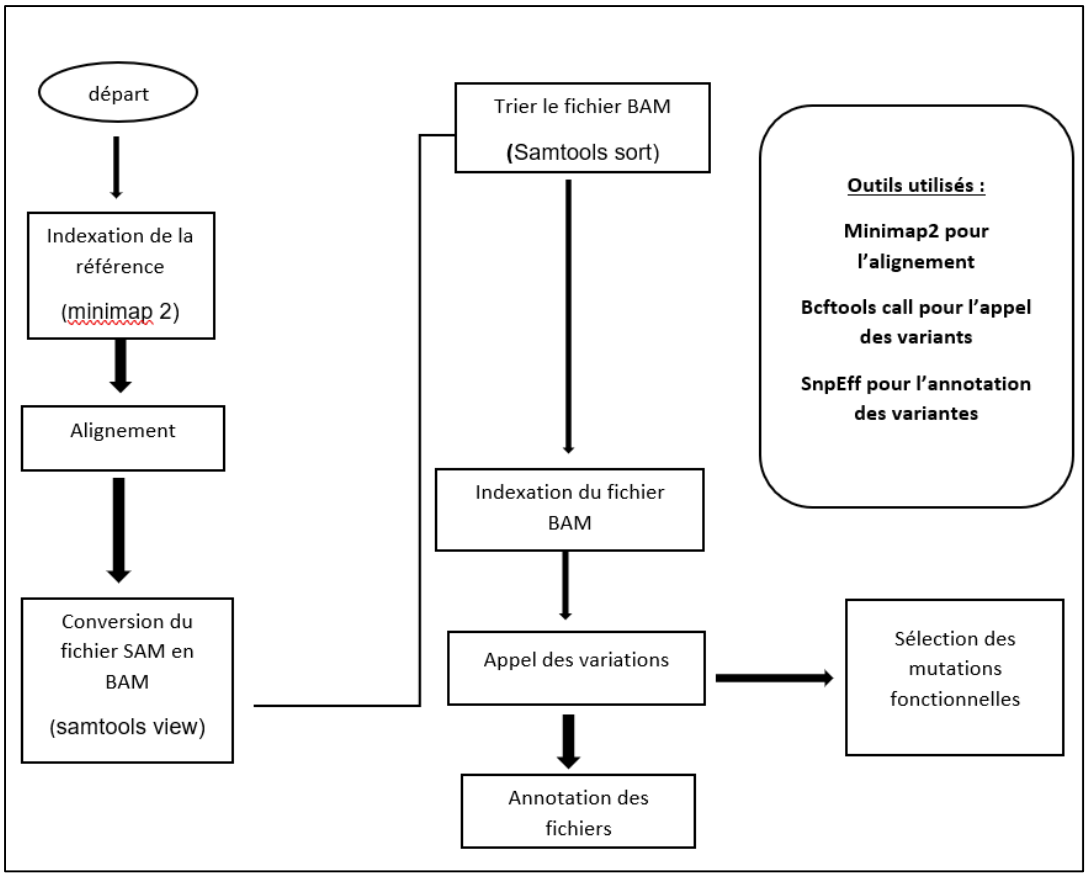
Annexe tableau 2: Mutations codantes et non synonymes identifiées au niveau de la séquence Omicron (FASTA).

pos	ref	alt	gene	mutation-nt	mutation-aa	peptide
2832	A	G	ORF1ab	2567A>G	K856R	NSP3
8393	G	A	ORF1ab	8128G>A	A2710T	NSP3
10029	C	T	ORF1ab	9764C>T	T3255I	NSP4
10449	C	A	ORF1ab	10184C>A	P3395H	NSP5
11537	A	G	ORF1ab	11272A>G	I3758V	NSP6
14408	C	T	ORF1ab	14144C>T	P4715L	NSP12
18163	A	G	ORF1ab	17899A>G	I5967V	NSP14
21762	C	T	S	200C>T	A67V	
21846	C	T	S	284C>T	T95I	
22578	G	A	S	1016G>A	G339D	
22673	T	C	S	1111T>C	S371P	
22674	C	T	S	1112C>T	S371F	
22679	T	C	S	1117T>C	S373P	
22686	C	T	S	1124C>T	S375F	
22813	G	T	S	1251G>T	K417N	
22882	T	G	S	1320T>G	N440K	
22898	G	A	S	1336G>A	G446S	
22992	G	A	S	1430G>A	S477N	
22995	C	A	S	1433C>A	T478K	
23013	A	C	S	1451A>C	E484A	
23040	A	G	S	1478A>G	Q493R	
23048	G	A	S	1486G>A	G496S	
23055	A	G	S	1493A>G	Q498R	
23063	A	T	S	1501A>T	N501Y	
23075	T	C	S	1513T>C	Y505H	
23202	C	A	S	1640C>A	T547K	
23403	A	G	S	1841A>G	D614G	
23525	C	T	S	1963C>T	H655Y	
23599	T	G	S	2037T>G	N679K	
23604	C	A	S	2042C>A	P681H	
23854	C	A	S	2292C>A	N764K	
23948	G	T	S	2386G>T	D796Y	
24130	C	A	S	2568C>A	N856K	
24424	A	T	S	2862A>T	Q954H	
24469	T	A	S	2907T>A	N969K	
24503	C	T	S	2941C>T	L981F	
26270	C	T	E	26C>T	T9I	
26530	A	G	M	8A>G	D3G	
26577	C	G	M	55C>G	Q19E	
26709	G	A	M	187G>A	A63T	
28311	C	T	N	38C>T	P13L	
28881	G	A	N	608G>A	R203K	
28883	G	C	N	610G>C	G204R	

Annexe Figures



Annexe Figure 1: Workflow de l'analyse génomique Illumina MiSeq.



Annexe Figure 2: Workflow de l'analyse génomique Ion Torrent.