

N° Ref :.....



Centre Universitaire Abdelhafid BOUSSOUF-Mila

Institut des Sciences et de la Technologie Département des Sciences de la Nature et de la Vie

Mémoire préparé En vue de l'obtention du diplôme de

Master

Domaine : Sciences de la Nature et de la Vie
Filière : Sciences Biologiques
Spécialité : Biochimie appliquée

Thème :

Etude phylogénétique du SARS-CoV-2 à la base du génome complet (2021)

Présenté par :

- KERROUCHE Meryem
- LAKHLEF Sara

Devant le jury composé de :

-Mme.HIMEUR Sara
-Mme.BELLATAR Hakima
-Mr. BOUBENDIR Abdelhafid

Maitre de conférences B
Maitre de conférences A
Maitre de conférences A

Président
Examinateur
Promoteur

Année Universitaire : 2020/2021



Remerciement

Au nom de Dieu le Miséricordieux, Nous tenons tout d'abord à remercier ALLAH le tout puissant et miséricordieux qui nous aide et nous donne la patience et le courage durant ces longues années d'étude, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

En second lieu, nous tenons à remercier notre encadreur **Dr BOUBENDIR Abd alhelhafid** pour son précieux conseil et son aide durant toute la période du travail et nous tenons d'abord à remercier très chaleureusement qui nous a permis de bénéficier de son encadrement.

En troisième lieu, nous présentons nos respects et nos remerciements aux membres du jury qui nous ont fait l'honneur d'évaluer ce travail président du jury **Mme.HIMEUR Sara** et **Mme.BELLATAR Hakima** examinateur, qui ont fait preuve d'intérêt et de disponibilité pour l'examen de notre mémoire.

Un grand merci à tous ceux qui ont contribué de près ou de loin à la réussite de cet ouvrage, qu'il soit matériel ou moral.



Dédicaces

*Un grand merci à Tout qu'ont collaboré de près ou de loin
à l'élaboration de ce travail*

□ *A ma très chère maman Farída*

*L'éducation et les principes que tu m'as inculqué m'ont
guidé tout au long de mon parcours. Tu t'es sacrifié pour
mon bien-être du début à la fin, je te remercie pour ton
énergie et tes prières. J'espère avoir été à la hauteur de
vos attentes. Que Dieu tout puissant vous préserve et vous
accorde santé, longue vie et bonheur.*

□ *A mon très cher père Abdalhamid*

*Papa, je te remercie d'avoir toujours cru en moi, cette
confiance a été ma source d'inspiration tout au long de
mes études. Tu t'es dévoué à donner le meilleur à tes
enfants en matière de connaissance et tu nous as
appris à nous battre jusqu'au bout pour réussir.
J'espère t'avoir honoré à travers ce travail. Dieu
Tout- Puissant vous préserve et vous donne santé,
bonheur et longue vie.*

□ *A mes Frères Dr yasser, Mohammed et Djaber*

Je vous souhaite un avenir plein de réussite et de sérénité.



Meryem

Dédicace

Avec l'expression de ma reconnaissance, je dédie ce modeste travail à ceux qui, quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

A l'homme, mon précieux offre du dieu, qui doit ma vie, ma réussite et tout mon respect, qui a su montrer à ses enfants que le travail est libérateur et qui n'a ménagé aucun effort pour l'aboutissement de ce travail, qu'il en soit récompensé pour ses sacrifices pour m'avoir enseigné le chemin du labeur et de la vérité. Que Dieu t'accorde une longue vie et une santé de fer avant que tu sois récompensé pour tous les sacrifices que tu as toujours consentis pour nous :

□ *Mon cher père Mourad* □

A la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureuse pour les sacrifices que tu consens pour rendre tes enfants heureux. Tu as enduré beaucoup de peine pour mon bien-être et à ma réussite. Reçois ceci en guise de ma reconnaissance et que Dieu te garde longtemps afin que tu puisses goûter aux arbres que tu as plantés qui n'a cessé de nourrir l'ambition de me conduire au succès. Reçois ma profonde gratitude. Daigne Dieu t'accorde une longue vie pour que bénéficie de tes efforts :

□ *Mon adorable mère Meriem* □

□ *A mes chers frères Amine et Abderraouf qui n'ont pas cessé de me conseiller, encourager et soutenir tout au long de mes études. Que Dieu les protège et leurs offre la chance et le bonheur.*

□ *A mes grands-mères, mes oncles et mes tantes. Que Dieu leur donne une longue et joyeuse vie. A tous les cousins, les voisins que j'ai connu jusqu'à maintenant. Merci pour leurs amours et leurs encouragements.*

□ *A tous mes amis et amies qui m'ont soutenu d'une manière ou d'une autre.*

□ *A tous ceux qui de près ou de loin, ont contribué à la réalisation de ce travail.*



Table des matières

Liste des abréviations	I
Liste des figures	III
Liste des tableaux	IV
Introduction	1

Partie I : Recherche Bibliographique

Chapitre 1 : Généralités	3
1. Définition	3
2. Historique	3
3. Taxonomie	4
4. Origine	4
5. Structure et organisation génétique	5
6. Mode de transmission	8
7. Facteurs influençant la propagation du SARS-CoV-2 dans le monde	9
7.1. Facteurs géodémographiques et environnementaux	9
7.2. Autres facteurs	12
8. Physiopathologie	13
Chapitre 2 : Biologie Moléculaire.....	15
1. Les ORFs	15
1.1 Définition	15
1.2. Les types ORFs	16
2. Les protéines non structurales	18
2.1. Les types NSPs	19
3. Les protéines accessoires	25
4. Les protéines structurales	25
4.1. Protéine M	25
4.2. Protéine E	26
4.3. La protéine N	27
4.4. La protéine S	28
5. Phylogénie du SARS-CoV-2	30

5.1. Phylogénie du génome complet	30
5.2. Phylogénie du génome partiel	33

Partie II : Recherche expérimentale

1. Matériel et méthodes	35
1.1. Présentation des bases de données	35
1.1.1. Présentation de NCBI	35
1.1.2. Présentation GISAID	35
1.2. Collecte des séquences génomiques à partir des bases de données	35
1.3. Etapes de la recherche sur NCBI	37
1.4. Les étapes de construction de l'arbre phylogénétique	40
1.4.1. Présentation de MEGA X	40
1.4.2. Alignement des séquences par muscle	41
1.4.3. Estimation des distances évolutives à l'aide de la distance par paires	46
1.4.4. La construction de l'arbre phylogénétique	48
2. Résultats et discussion	50
Conclusion	55
Liste de références.....	56

Liste d'abréviations

ARN : Acide Ribo Nucléique

ARNm : Acide Ribo Nucléique messenger

ADN : Acide désoxyribonucléique

SARS-CoV-2 : Coronavirus du Syndrome Respiratoire Aigu Sévère 2

COVID19 : Coronavirus Disease 2019

nCoV : nouveau CoronaVirus

OMS : Organisation Mondial de la Santé

USPPI : Urgence de Santé publique de Porté Internationale

MERS-CoV : Coronavirus du syndrome respiratoire du Moyen-Orient

ACE2 : Enzyme de Conversion l'Angiotensine 2

ORF : Open Reading Frame

NSP : Non Structurel Protéine

TMPSS2 : Transmembrane protéase serine subfamily member 2

UTR : Untranslated Transcribed Region

RBD : Receptor Binding Domaine

ER : reticulum endoplasmique

2019nCoVVR : Novel CoronaVirus Resource

NCBI : Centre for Biotechnology Information

GISAID : Initiative mondiale sur le partage de toutes les données sur la grippe

PP : Polyprotéines

RdRp : ARN Polymérase ARN dépendant

PHB : Prohibitin

NF- κ B : Facteur nucléaire Kappa B

IRF : L'interféron de l'hôte régulateur facteur 3

3CL-PRO : Protéase de type chymotrypsine

Mpro : Main protéase

DMV : Virus Double Membrane

ZBD : Zinc Binding Domain

N7MTase : N7 Méthyl transférase

IFN : Interféron

TM : Transmembranaire

NTD : Domaine N Terminal

CTD : Domaine C Terminal

FP : Fusion Peptide

HR1 : Heptapeptide séquence de répétition

NA : Numéro d'Accès

MEGA : Molecular Evolutionary Genetics Analysis

NJ : Neighbor-Joining

Liste de figures

Figure 1. Transmission de coronavirus de l'origine jusqu'à l'homme (Machhi et al., 2020)	5
Figure 2. Structure du SARS-CoV-2 (Kumar et al., 2020)	7
Figure 3. Organisation génétique de SARS-CoV-2 (Segondy, 2020)	8
Figure 4. Mode de transmission du SARS-CoV-2 (Harrisonne et al., 2020).....	9
Figure 5. Le rôle de facteurs environnementaux dans la propagation du SARS-CoV-2 (Eslami et Mahrokh, 2020)	11
Figure 6. Le cycle de vie du SARS-CoV-2 dans les cellules hôtes (Shereen et al., 2020)	14
Figure 7. Les gènes de génome de SARS-CoV-2 (Jungereis et al., 2020)	15
Figure 8. Structure de protéine M (Thomas, 2020)	26
Figure 9. Structure de protéine E (Thomas, 2020).....	27
Figure 10. Structure de protéine N (Thomas, 2020).....	28
Figure 11. Structure de protéine S (Huang et al., 2020)	30
Figure 12. Structure 3D protéine S (PDB, 2021).....	30
Figure13. Réseau phylogénétique de 160 génomes du SARS-CoV-2 (Forster et al., 2020)	31
Figure 14. Estimation de la phylogénie du maximum de vraisemblance du coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2). Les couleurs indiquent les différents emplacements d'échantillonnage. L'arbre est enraciné au milieu. Les résultats des analyses phylogénétiques de vraisemblance maximale de « dataset-70 » sont affichés (Li et al., 2020)	32
Figure 15 . Arbre phylogénétique pour 322 nucléotides du gène nsp-2 du SARS-CoV-2 en utilisant la méthode de jointure voisine et 1000 bootstrap, le triangle rouge montre la séquence de référence pour le SARS-CoV-2 et le triangle bleu indique Iranien isole. (Pour l'interprétation des références à la couleur dans cette légende de la figure, le lecteur est renvoyé à la version Web de cet article) (Tabibzadeh et al., 2020)	34
Figure 16. La fenêtre principale de MEGA X.....	41
Figure 17. Arbre phylogénétique de génome complet du SARS-CoV-2 de différente région géographique. Les isolats sont représentés avec, le pays/ le numéro d'accès d'origine et l'année d'isolement	50

Liste de tableaux

Tableau 1. Classification taxonomique du SARS-CoV-2 (Gligorov et al., 2020 ; Gorbalenya et al., 2020)	4
Tableau 2. L'expression des gènes de SARS-CoV-2 (NC045512.2) (Yoshimoto et al., 2020)	18
Tableau 3. Les protéines non structurelle et leur fonction moléculaire (Naqvi et al., 2020 ; Yoshimoto et al., 2020)	24
Tableau 4. Les séquences génomiques à partir des bases de données.....	36
Tableau 5. Répartition des isolats dans les clades	51
Tableau 6. Les valeurs de distance évolutionnaire entre les pays	52

Introduction

Introduction

Fin décembre 2019, une épidémie d'une maladie inconnue appelée pneumonie de cause inconnue s'est produite à Wuhan, dans la province du Hubei en Chine. Quelques jours plus tard, l'agent causal de cette mystérieuse pneumonie a été identifié comme un nouveau coronavirus (nCoV) par plusieurs laboratoires indépendants. Le virus en cause a été temporairement nommé coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) (He et al., 2020).

Le SARS-CoV-2 est un nouveau coronavirus humain. Il a émergé en Chine fin 2019 et est responsable de la pandémie mondiale de Covid19 qui a causé plus de 540 000 décès en six mois. La compréhension de l'origine de ce virus est une question importante et il est nécessaire de déterminer les mécanismes de sa dissémination afin de pouvoir se prémunir de nouvelles épidémies. L'analyse phylogénétique des séquences et les relations structure-fonction des protéines de coronavirus éclairent les connaissances actuellement disponibles sur les différents scénarios pour rendre compte de l'origine naturelle ou synthétique du virus (Sallard et al., 2020).

L'entrée virale repose sur une interaction fine entre le virion et la cellule hôte. L'infection est initiée par l'interaction de la particule virale avec des protéines spécifiques sur la surface cellulaire et peut avoir été transmis à l'homme par un hôte animal intermédiaire (Li et al., 2020).

L'épidémie de COVID-19 est devenue une menace sanitaire mondiale, elle provoque plusieurs maladies impliquant les systèmes respiratoire, entérique, hépatique et neurologique avec une gravité variable chez les humains et les animaux. Les principaux symptômes de COVID-19 comprenaient la fièvre, la fatigue et la toux, Il existe certains aspects distincts et superposés de la pathologie et de la pathogenèse des CoV qui provoquent des maladies graves chez l'homme. De nombreuses publications ont rapporté les caractéristiques cliniques, virologiques, pathologiques et radiologiques de COVID-19, mais l'examen complet est peu nombreux (He et al., 2020).

Notre étude est subdivisée en deux parties essentielles, la première partie présente une synthèse bibliographique dans laquelle nous apportons un premier chapitre qui est consacré à l'étude du coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2), un second chapitre expose la biologie moléculaire du virus.

La deuxième partie expérimentale est réservée à l'étude phylogénétique du génome complet du SARS-CoV-2. Cette partie vise à récolter les séquences nucléiques du génome complet du SARS-CoV-2 disponibles dans les bases de données sur internet, les isolats sélectionnés sont issus de différentes géographies dans le monde, parmi eux des isolats Algériens sont ciblés. Ensuite vient la construction de la matrice des distances évolutives et les dendrogrammes. L'outil bioinformatique utilisé dans la présente étude est le logiciel MEGA X.

Partie I
Recherche
Bibliographique

Chapitre 1 : Généralités

Chapitre 1 : Généralités

1. Définition

Les coronavirus (du latin corona) sont une très large famille de virus qui doivent leur nom au fait qu'ils semblent dotés d'une couronne (Didier, 2020), ils sont des virus à ARN simples brins responsables d'affections respiratoires dans la population générale (Van Nho et al., 2020).

Ils contiennent les plus grands génomes de tous les virus à ARN, enveloppés, à sens positif et ont un seul brin (He et al., 2020). Les coronavirus appartiennent à la sous-famille des Coronavirinae dans la famille des coronaviridae, et cette sous-famille comprend quatre genres : α -coronavirus, β -coronavirus, γ -coronavirus et δ -coronavirus. Le SARS-CoV-2 appartient à la famille des coronavirus, genres β -coronavirus (Xu et al., 2020).

Le SARS-CoV-2 provoque une infection appelée maladie à coronavirus 2019 (COVID-19), il est la septième forme de la famille des coronavirus infectant l'espèce humaine (Van Nho et al., 2020). Il s'agit d'un virus enveloppé dont le génome est constitué d'un acide ribonucléique (ARN) simple brin de polarité positive, d'une longueur d'environ 30 kb (OMS, 2021).

2. Historique

Depuis décembre 2019, le monde entier vit au rythme d'une crise sanitaire grave provoquée par l'émergence d'un nouveau coronavirus (Yahia et al., 2020). Le 5 janvier 2020, l'OMS a annoncé que l'épidémie de pneumonie était due à un nouveau virus originaire du marché Huanan Sea Food à Wuhan, en Chine, le 12 janvier 2020, le virus a été confirmé comme un nouveau coronavirus (nCoV). Le 30 janvier 2020, l'OMS a déclaré que le 2019-nCoV était une urgence de santé publique de portée internationale (USPPI). Le 11 février 2020, le Comité international de taxonomie des virus a officiellement renommé le nouveau coronavirus (anciennement 2019-nCoV) coronavirus du syndrome respiratoire aigu sévère 2 (SARS-CoV-2) responsable de l'épidémie actuelle de COVID-19 (Li et al., 2020), pour « Coronavirus disease 2019 », par l'Organisation Mondiale de Santé (OMS) (Ge et al., 2020).

Il convient de noter que la famille des Coronaviridae comprend non seulement le SARS-CoV-2, mais aussi le SARS-CoV, le coronavirus du syndrome respiratoire du Moyen-Orient (MERS-CoV) et les virus du rhume (ex., 229E, OC43, NL63 et HKU1) (Li et al., 2020).

3. Taxonomie

SARS-CoV-2 est un virus qui appartient à la famille des Coronavirus (CoV). Il est par ailleurs le second virus de cette famille identifié comme responsable d'épidémies de COVID 19 (Gligorov et al., 2020), il appartient à l'ordre des Nidovirales (Tableau 1) dans la famille des Coronaviridae, sous-famille des Coronavirinae et genres β -coronavirus (Xu et al., 2020).

Tableau 1. Classification taxonomique du SARS-CoV-2 (Gligorov et al., 2020 ; Gorbalenya et al., 2020).

Domaine	Riboviria
Ordre	Nidovirales
Sous-ordre	Cornidovirineae
Famille	Coronaviridae
Sous-famille	Orthocoronavirinae
Genre	Betacoronavirus
Sous-genre	Sarbecovirus
Espèce	Coronavirus liés au syndrome respiratoire aigu sévère (SARS-CoV)
Souche	Coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2)

4. Origine

Des coronavirus humains endémiques, trois nouveaux coronavirus d'origine zoonotique ont émergé dans la population humaine ces vingt dernières années. Il s'agit du SARS-CoV apparu en 2003, du MERS-CoV apparu en 2012, et du SARS-CoV-2 apparu en 2019 (Segondy, 2020).

Le virus provenait à l'origine de chauves-souris et été transmis au fil du temps à d'autres animaux hôtes et finalement aux humains (Figure 1) (Ludwig et Zarbock, 2020; Machhi et al 2020). Les lieux de vie des chauve-souris étant éloignés des communautés humaines donc le passage interespèces a probablement nécessité un hôte intermédiaire (Bonny et al., 2020). Les virus de cette famille infectent plus de 500 espèces de chiroptères (ordre de mammifères comprenant plus de 1200 espèces de chauves-souris) (Sallard et al., 2020).

L'analyse phylogénétique a révélé que le SARS-CoV-2 est étroitement lié (similitude de 88 à 89%) à deux coronavirus de type SARS dérivés de chauves-souris, à savoir chauve-souris-SL-CoVZC45, et bat-SL-CoVZXC21, et elle indique que le SARS-CoV-2 est similaire au coronavirus circulant chez *Rhinolophus* (fer à cheval), avec 98,7% de similitude nucléotidique avec le gène de l'ARN polymérase ARN-dépendante partielle (RdRp) de la souche de coronavirus de chauve-souris BtCoV / 4991 (GenBank KP876546, séquence de 370 pb de RdRp) et 87,9% de similitude nucléotidique avec la souche de coronavirus de chauve-souris bat-SL-CoVZC45 et bat-SL-CoVZXC21 (Sallard et al., 2020).

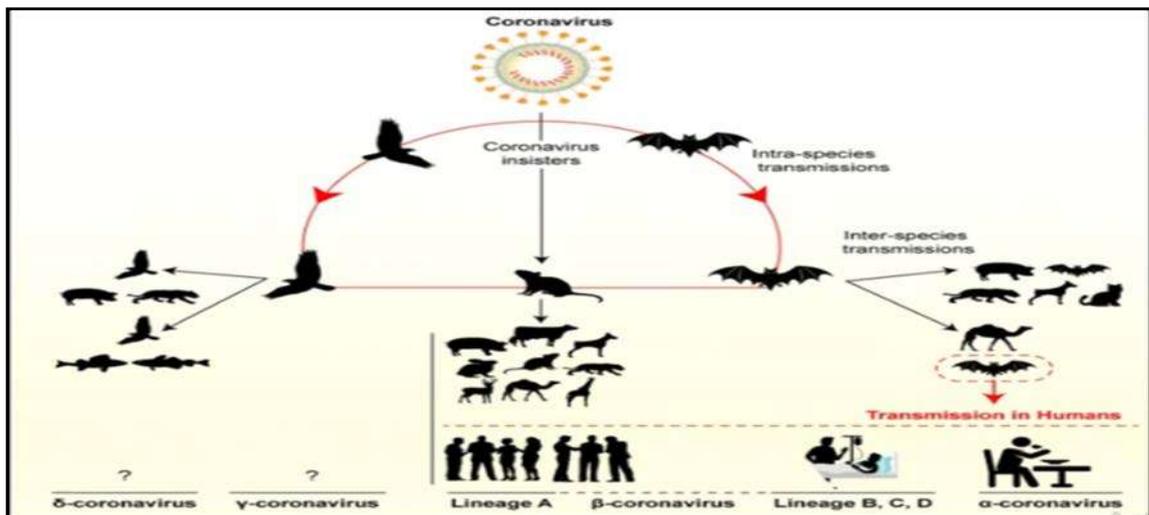


Figure 1. Transmission de coronavirus de l'origine jusqu'à l'homme (Machhi et al., 2020).

5. Structure et organisation génétique

Les CoV possèdent un génome comparativement volumineux approchant 30 kb (Kandeel et al., 2020), ce sont des virus sphériques et pléomorphes avec diamètre 80 à 160 nm (Machhi et al., 2020). Le séquençage génétique permet la lecture des génomes viraux

étant donné que chaque agent pathogène possède une séquence génomique unique, cette méthode peut être utilisée pour identifier de nouveaux agents pathogènes (comme dans le cas du SARS-CoV-2). Le génome du SARS-CoV-2 code pour des protéines non structurales, quatre protéines structurales (spicule ou « spike » (S), enveloppe (E), membrane (M), nucléocapside (N) et plusieurs protéines accessoires putatives (OMS, 2020).

La protéine S est une grande protéine transmembranaire de type I (Belouzard., 2012), c'est une glycoprotéine de la surface des CoV peut se fixer au récepteur cellulaire, elle forme des homotrimères qui émergent à la surface du virus et facilite leur liaison aux cellules hôtes par attraction avec l'enzyme de conversion de l'angiotensine 2 ACE2 se trouve dans les voies respiratoires inférieures de l'homme (Ludwig et Zarbock, 2020).

La nucléocapside connue sous le nom de protéine N est le composant structurel de la localisation du CoV dans la région du réticulum endoplasmique-Golgi qui est structurellement liée au matériel d'acide nucléique du virus. Parce que la protéine est liée à l'ARN, pour former une ribonucléoprotéine, elle s'implique dans les processus liés au génome viral, au cycle de réplication virale, et la réponse cellulaire des cellules hôtes aux infections virales. La protéine N est également fortement phosphorylée et elle est suggérée de conduire aux changements structuraux améliorant l'affinité pour l'ARN viral (Astuti, 2020).

Une autre partie importante de ce virus est la membrane ou protéine M, qui est la protéine la plus structurellement structurée et joue un rôle dans la détermination de la forme de l'enveloppe du virus. Cette protéine peut se lier à toutes les autres protéines structurales. La protéine aide à stabiliser les nucléocapsides ou les protéines N et favorise l'achèvement de l'assemblage viral en stabilisant la protéine N-ARN complexe, à l'intérieur du virion interne (Astuti, 2020).

Le dernier composant est l'enveloppe ou protéine E qui est la plus petite protéine de la structure SARS-CoV qui joue un rôle dans la production et la maturation de ce virus (Astuti, 2020), elle aide à l'assemblage des virions et comprend les actions des canaux ioniques (Figure 2) (Kumer et al., 2020 ; OMS, 2021).

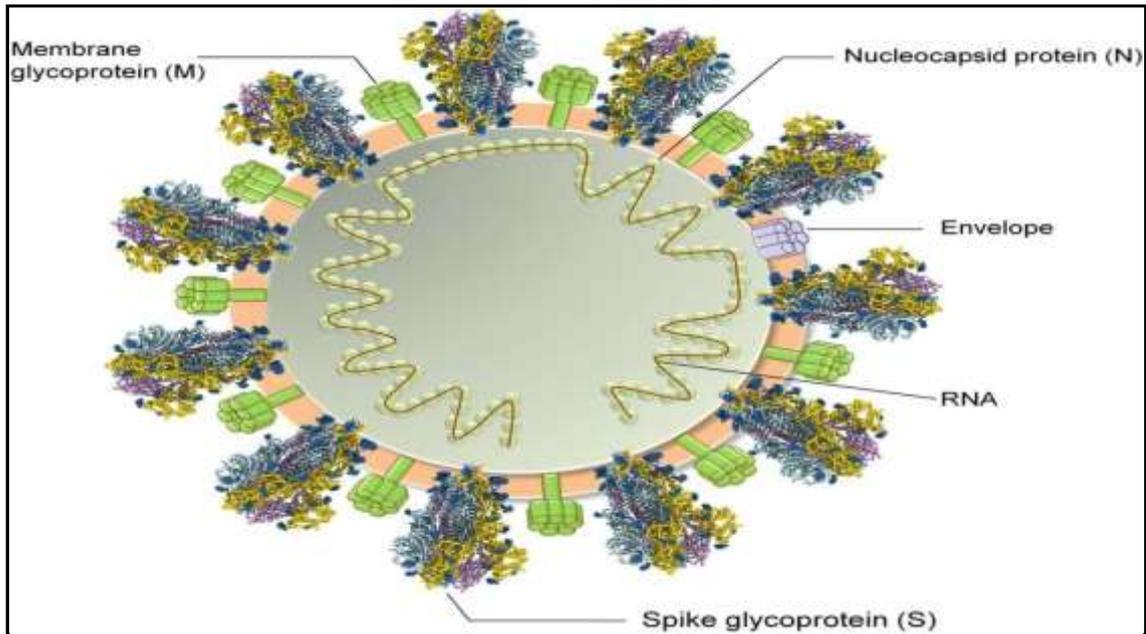


Figure 2. Structure du virus SARS-CoV-2 (Kumar et al., 2020).

Tous les CoV ont en commun que leur génome est dans la forme d'acide ribonucléique simple brin (ARN) avec polarité positive, ce qui signifie que la séquence de base de l'ARN est dans l'orientation 5' → 3' et correspond au dernier ARN messager (ARNm) (Luwdig et Zarbock, 2020).

A la différence de tous les autres virus à ARN connus le génome du CoV est le plus grand génome ARN d'une longueur de 26,4 à 31,7 kilobases (Luwdig et Zarbock, 2020), et comprend 6 à 11 cadres de lecture ouverts (ORF) codant pour 9680 polyprotéines d'acides aminés (Kumar et al., 2020).

Le génome du SARS-CoV-2 comprend 2 régions non traduites (UTR) à l'extrémité 5' (265 nucléotides) et à l'extrémité 3' (358 nucléotides) et une région codante divisée en plusieurs ORF. Le premier ORF comprend environ 67% du génome qui code pour 16 protéines non structurales (NSPs), tandis que les ORF restants codent pour des protéines accessoires et structurelle (Kumar et al., 2020 ; Wu et al., 2020).

La variation de séquence entre le SARS-CoV-2 et le SARS-CoV n'a révélé aucune différence significative dans les ORFs et les NSPs (Kumar et al., 2020).

Le NSP comprend deux cystéine protéases virales, y compris la protéase de type papaine (NSP3), la protéase de type chymotrypsine, de type 3C ou principale (NSP5), l'ARN polymérase dépendante de l'ARN (NSP12), l'hélicase (NSP13) et d'autres probables

d'être impliqué dans la transcription et la réplication du SRAS-CoV-2 (Figure 3) (Kumar et al., 2020 ; Segondy, 2020).

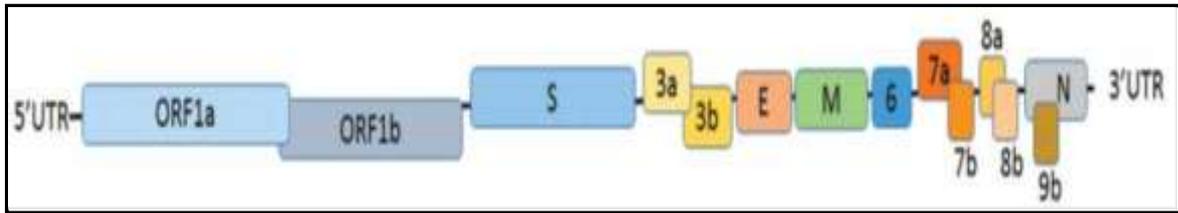


Figure 3 : Organisation génomique du SARS-CoV-2 (Segondy, 2020).

6. Mode de transmission

La transmission de SARS-CoV-2 de l'animal à l'homme était présumée être le principal mécanisme. Néanmoins, les cas ultérieurs n'étaient pas associés à ce mécanisme d'exposition et il a été conclu que le virus pouvait être transmis d'humain à humain (Gligorov et al., 2020).

Le SARS-CoV-2 est un virus pneumotrope qui se propage d'une personne à l'autre principalement par les sécrétions respiratoires (Triggle et al., 2021), y compris les gouttelettes chargées de particules virales pourraient infecter soit par contact direct avec une muqueuse (transmission directe) soit par contact avec une surface infectée par les muqueuses nasales, buccales ou conjonctivales (transmission indirecte). Elles peuvent être projetées à plusieurs mètres de distance mais ne persistent pas dans l'air. Bien que le virus puisse survivre au moins trois heures dans l'air, l'ARN viral a également été détecté dans les selles et le sang des patients infectés (Figure 4) (Bonny et al., 2020).

Dans les environnements intérieurs avec une ventilation inadéquat le SARS-CoV-2 reste très contagieux (Triggle et al., 2021).

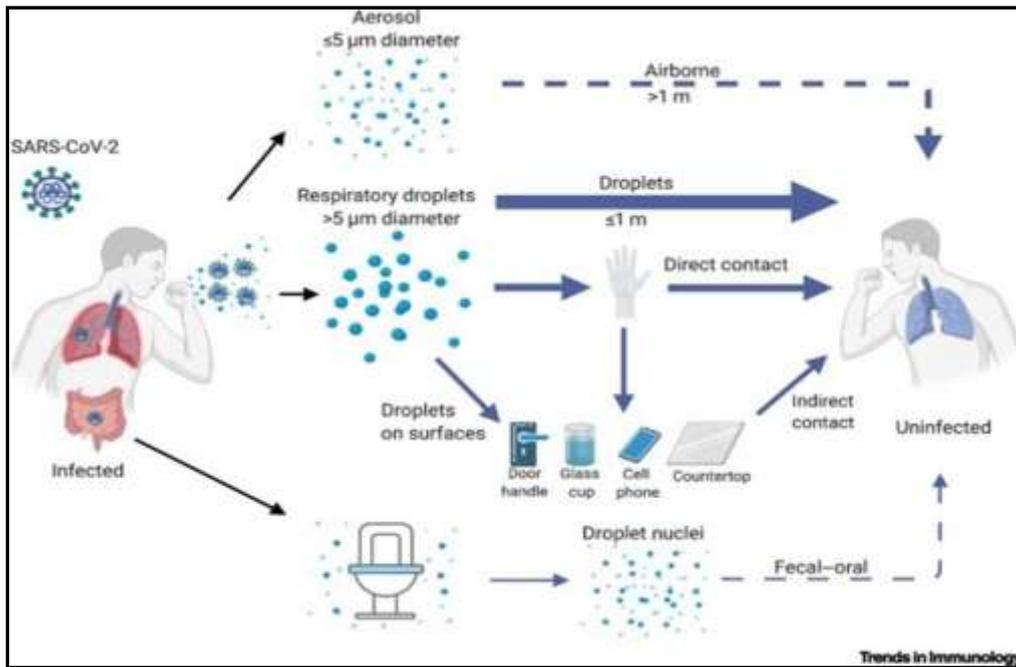


Figure 4. Mode de transmission de ARS-CoV-2 (Harrisonne et al., 2020).

7. Facteurs influençant la propagation de SARS-CoV-2 dans le monde

L'épidémie actuelle de la nouvelle maladie à coronavirus 2019 (COVID-19) dans plusieurs pays est devenue une menace sérieuse pour la santé des personnes dans le monde. La transmission interhumaine du SARS-CoV-2 se produit le plus souvent lorsque les personnes sont au stade d'incubation de la maladie ou sont porteuses et ne présentent aucun symptôme. Par conséquent, il y a des facteurs et des conditions environnementales tels que température humidité, transmission par les aliments et les emballages des aliments, l'air qui jouent un rôle dans la transmission du COVID-19 (Figure 5) (Eslami et Mahrokh, 2020).

7.1. Facteurs géodémographiques et environnementaux

- **L'eau et les eaux usées**

Dans l'étude d'Eslami et Mahrokh (2020), le rapport de la présence du virus SARS-CoV-2 dans les eaux usées a été confirmé. La présence du virus dans les eaux usées, même lorsque la prévalence du virus est faible, indique que le SARS-CoV-2 peut survivre dans les eaux usées. Certaines études ont rapporté que le coronavirus peut rester dans l'eau ou les sources d'eaux usées pendant des jours ou des semaines. En raison des preuves de la survie du virus dans les environnements aquatiques et d'égouts, sa présence dans les

ressources en eau dépend de facteurs essentiels tels que la température, la lumière du soleil et la présence de composés organiques auxquels le virus peut s'adsorber et se protéger contre la lumière du soleil. Selon le dernier rapport de l'OMS, il n'y a aucune preuve que le coronavirus de type humain ait été transmis par l'eau potable contaminée (Eslami et Mahrokh, 2020).

- **Les insectes**

Certaines études ont associé des insectes tels que les coléoptères et les insectes domestiques, qui sont les principaux vecteurs mécaniques d'agents pathogènes, par contact avec des surfaces contaminées et des sécrétions de patients impliquées dans la transmission. L'excrétion du SARS-CoV-2 par les selles a été confirmée chez certains patients. Par conséquent, l'alimentation des insectes domestiques et des coléoptères à partir des selles et sa transmission mécanique peut jouer un rôle important dans la transmission de la maladie. Ils peuvent transmettre plus de 100 agents pathogènes par les jambes, les poils, la bouche, les selles et les vomissures (Eslami et Mahrokh, 2020).

- **L'âge**

La population particulièrement jeune, serait à l'origine d'une propagation rapide de l'infection. En effet, la plupart des individus jeunes seraient asymptomatiques ou présenteraient des formes bénignes, et passeraient probablement inaperçus, avec un risque d'infecter plus de personnes que ceux qui sont symptomatiques (Hoummadi et al., 2020)

- **Facteurs climatiques**

Depuis plus de cinquante ans, la relation entre les conditions météorologiques et les maladies est étudiée. De telles études sont importantes pour prédire la propagation des maladies virales, en particulier si cela conduit à des pandémies comme dans le cas de la covid-19, afin d'aider les décisions dans les politiques de santé publique au niveau mondial (Demongeot et al., 2020).

Une étude a observé la relation entre le nombre de cas quotidiens positifs de SRAS-CoV-2 avec trois facteurs environnementaux : humidité relative maximale, température maximale et la vitesse maximale du vent (Eslami et Mahrokh, 2020).

La température est l'un des principaux facteurs environnementaux liés aux variations saisonnières et régionales. La température environnementale a un effet fort et régulateur sur la durée de survie et la transmission du SARS-CoV-2 par le biais de gouttelettes et

d'objets non vivants. Plusieurs études précédentes avaient signalé un effet régulateur de la température sur la survie et la transmission d'autres coronavirus, ces études ont rapporté que la température environnementale optimale pour les cas de SARS se situait entre 16 °C et 28 °C. Environ 90 % des cas et des décès dans le Hubei, la Lombardie, Madrid, Londres, New York et le Wisconsin ont été détectés entre 7,5 °C et 14 °C de température moyenne par jour. Environ 90% des cas et des décès ont été détectés à Western Cape, Sau Paulo, Sydney, Dhaka et Maharashtra entre 16,5 °C et 30°C de température moyenne par jour. Ces résultats étaient en bon accord avec les études précédentes, car une fréquence plus élevée des cas et des décès ont été détectés en dessous de 20 °C dans le monde (Sharif et al., 2021).

L'humidité relative et la vitesse du vent jouent également un rôle important dans la propagation saisonnière et régionale des coronavirus. La majorité des cas et des décès à 65% ont été enregistré à 80 % d'humidité relative moyenne par jour, ce qui est proche de l'optimum pour que les noyaux de gouttelettes survivent plus longtemps dans l'environnement. Pour le test de Spearman, cette étude rapporte que l'humidité relative moyenne et la vitesse du vent sont significatives pour les cas et les décès dans différentes régions du monde (Sharif et al., 2021).

Il y'a un autre facteur environnemental qui est l'indice UV, la majorité des cas et des décès ont été détectés à un indice UV moyen de 3 à 6 par jour dans les régions d'étude. Cependant, la transmission du SARS-CoV-2 a été détectée à des indices UV moyens très élevés et extrêmes de 9,3 et 11,7 par jour à Dhaka et Maharashtra, respectivement (Sharif et al., 2021).

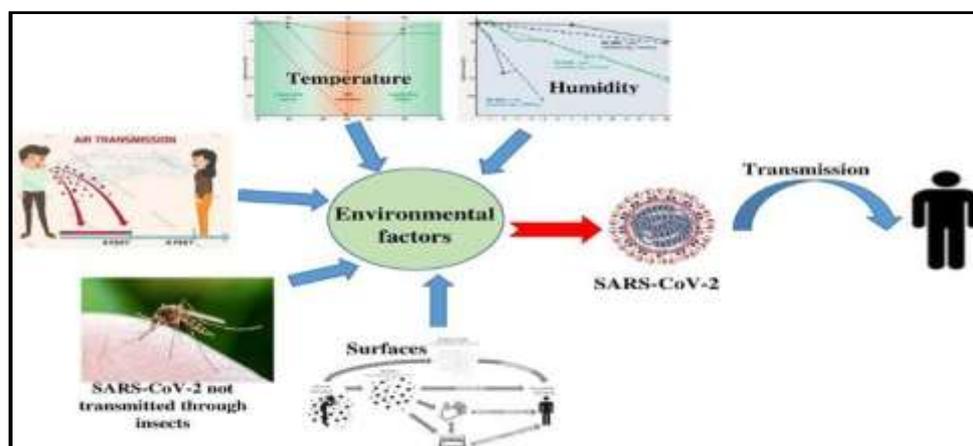


Figure 5. Le rôle de facteurs environnementaux dans la propagation du SARS-CoV-2 (Eslami et Mahrokh, 2020).

7.2. Autres facteurs

- **L'immigration**

La propagation du COVID-19 depuis la Chine a été clairement liée à ceux qui voyagent de Wuhan dans la province du Hubei en Chine centrale. Par conséquent, il est important de comprendre la densité/le volume de voyage des passagers transportés ainsi que les itinéraires au départ de Wuhan via les principaux centres de transport aérien régionaux connectés à travers la Chine. Un modèle a été développé sur l'intensité de la migration et des voyages qui peut expliquer l'épidémie et la propagation du COVID-19 depuis son apparition fin 2019. La présence de populations de migrants d'origine chinoise et de stock d'immigrants en Chine sont des indicateurs utiles dans la prédiction de la propagation de l'épidémie dans le monde en cas d'interaction avec plusieurs autres facteurs macro. La surveillance des données sur le stock d'immigrants et les données sur le volume de déplacements basées sur les couloirs de mobilité humaine (c'est-à-dire les origines et les destinations), les pays auraient pu être mieux préparés et prendre des mesures précoces pour contenir la propagation de COVID-19 (Sirkeci et Mustafa, 2020).

Nous avons observé un taux de notification COVID-19 plus élevé chez les immigrants par rapport aux non-immigrants et un taux d'hospitalisation beaucoup plus élevé, avec des différences majeures entre les différents groupes d'immigrants. Les immigrants somaliens, pakistanais et irakiens avaient des taux particulièrement élevés (Indseth et al., 2020).

- **Tourisme**

Le COVID-19 a émergé dans un monde étroitement connecté par la population locale et internationale en mouvements, avec plus de personnes se déplaçant pour le travail, l'éducation et des raisons familiales, le tourisme et survie que jamais dans le passé. Des mouvements de population intenses, en particulier des touristes et les travailleurs d'affaires, ont été un facteur clé de la propagation mondiale de l'épidémie (Indseth et al., 2020).

- **Les échanges commerciaux**

Des études montrent clairement l'existence de solides relations commerciales/économiques entre la Chine et l'Italie du Nord. La corrélation entre les cas d'infection et l'importation commerciale et données d'exportation a été quantifiée : cela a donné 0,99 en considérant les cas d'infection par rapport aux données d'importation, et 0,97 en considérant les cas d'infection par rapport aux données d'exportation. Ceci couplé avec la considération que la Lombardie est la plus région italienne peuplée, soutient fortement

l'idée qu'économique/commerciale relations peuvent être considérées comme le facteur le plus probable contribuant à la phase initiale de diffusion de l'infection en Italie du Nord. Les relations commerciales peuvent être utiles paramètre représentant les modèles de mobilité à travers les zones étudiées. En particulier, c'est probable qu'une mesure appropriée des mécanismes de diffusion initiale du virus d'humain à humain pourraient être trouvés dans le commerce Italie-Chine (mobilité externe). Il est évident que d'autres facteurs peuvent avoir contribué à affecter la dynamique de transmission et à accélérer la propagation du COVID-19 comme la mobilité interne des personnes, les conditions environnementales la possibilité de l'interaction du virus avec l'environnement, mais le facteur économique/commercial semble être la raison la plus plausible pouvant justifier la phase initiale de diffusion du COVID19 (Bontempi, 2020).

8. Physiopathologie

La physiopathologie et les mécanismes de virulence du SARS-CoV-2 ont des liens avec la fonction des NSPs et protéines structurales (Rajnik et al., 2021).

La principale voie d'entrée du SARS-CoV-2 passe par les voies respiratoires supérieures ou les muqueuses faciales (Triggle et al., 2021).

Parmi les fonctions des protéines structurales, l'enveloppe joue un rôle crucial dans la pathogénicité du virus car elle favorise l'assemblage et la libération virale (Rajnik et al., 2021). La protéine Spike est en réalité composée de 2 parties : S1 un domaine de liaison au récepteur (RBD) qui assure le contact direct avec le principal récepteur cellulaire (ACE2), et S2 est importante pour la fusion de la membrane du virus avec celle de la cellule, c'est-à-dire, le premier temps indispensable à la pénétration intracellulaire du virus, et un site de clivage polybasique S1 / S2 qui est clivée protéolytiquement par la cathepsine L cellulaire et la protéase transmembranaire sérine 2 (TMPRSS2) ce dernier facilite l'entrée virale au niveau de la membrane plasmique donc, la seule présence de l'ACE2 sur la membrane d'une cellule ne permet pas au virus d'entrer ainsi la NSP est capable de bloquer la réponse immunitaire innée de l'hôte (Harrison et al., 2020 ; Rousseau et al., 2020 ; Rajnik et al., 2021).

Le récepteur ACE2 est exprimé à 80% au niveau des cellules alvéolaires pulmonaires de type II, mais aussi au niveau des muqueuses nasales, de l'arbre respiratoire supérieur, de

l'endothélium vasculaire, ainsi que des tissus cardiaques, rénaux et digestifs. La sérine protéase TMPRSS2, très exprimée par les cellules endothéliales des tractus respiratoires et digestifs, semble aussi jouer un rôle important comme co-récepteur du virus (Couvreur et al., 2021).

Une fois que le génome est libéré dans le cytosol hôte, les gènes de cette génome (ORF1a et ORF1b) sont traduits en polyprotéines de réplicase virale (pp1a et pp1ab), qui sont clivées en petite produit par des protéinase virale (NSPs) ceux-ci forment l'ARN dépendant de l'ARN polymérase (Figure 5) (Harrison et al., 2020 ; Shereen et al., 2020).

L'ARN génomique sens négatif (-) est synthétisé et utilisé comme matrice pour former un ARN sens subgénomique ou génomique positif (+). L'ARN viral et la protéine structurale de la nucléocapside (N) sont répliqués, transcrits ou synthétisés dans le cytoplasme, tandis que d'autres protéines structurales virales, y compris S, membrane (M) et enveloppe (E), sont transcrites puis traduites dans le réticulum endoplasmique (ER) et transporté au Golgi. Le complexe ARN-N viral et les protéines S, M et E sont ensuite assemblés dans le compartiment intermédiaire ER-Golgi pour former un virion mature, puis libérés des cellules hôtes par exocytose (Jiang et al., 2020).

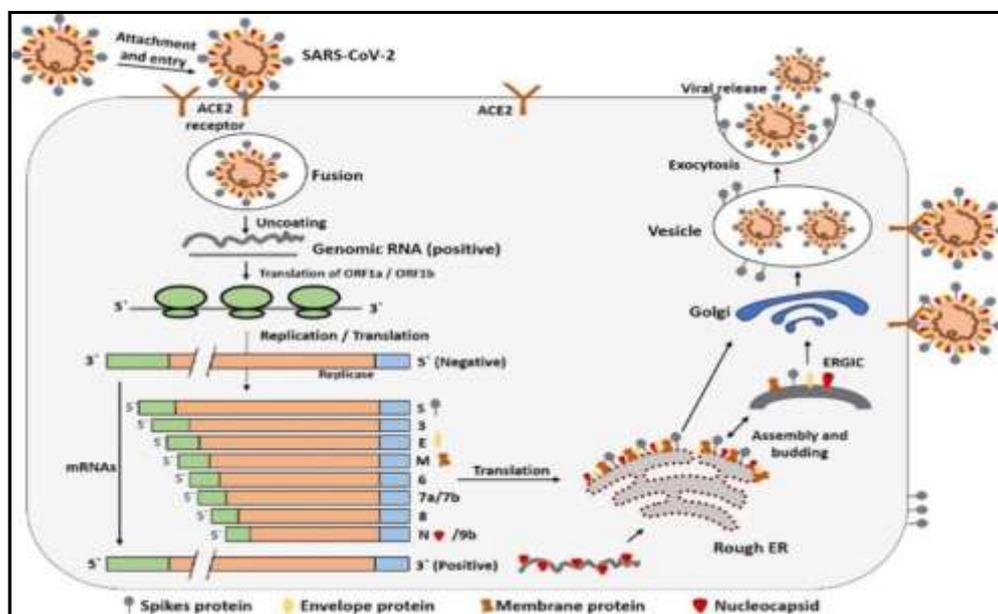


Figure 6. Le cycle de vie du SARS-CoV-2 dans les cellules hôtes (Shereen et al., 2020).

Chapitre II :

Biologie Moléculaire

Chapitre 2 : Biologie Moléculaire

1. Les ORFs

1.1 Définition

Tout comme l'ADN génomique présente dans le noyau de chacune de nos cellules, l'ARN génomique des coronavirus porte l'information génétique indispensable à la production de nouveaux virions (Juckel et al., 2020).

Les séquences génomiques complètes du SARS-CoV-2 ont été obtenues à partir de 2019 Novel Coronavirus Resource (2019nCoV) et de deux bases de données, dont le National Center for Biotechnology Information (NCBI) et l'Initiative mondiale sur le partage de toutes les données sur la grippe (GISAID) (Xu et al., 2020).

Le génome nouvellement séquencé du SARS-CoV-2 a été soumis dans la base de données génomique du NCBI (NC_045512.2) (Miao et al., 2020 ; Naqvi et al., 2020).

Le génome de l'ARN est coiffé à l'extrémité 5' et polyadénylé à l'extrémité 3' (Miao et al., 2020). Le SARS CoV-2 contient au moins six cadres de lecture ouverts (ORF) et de nombreux autres gènes accessoires comme d'autres CoV. Les deux tiers terminaux 5' du génome contiennent deux cadres de lecture ouverts (ORF), ORF1 et ORF2 qui code pour deux polyprotéines, pp1a et pp1ab, qui est ensuite clivée en 11 et 16 protéines, respectivement. Ces 16 protéines matures sont responsables de plusieurs fonctions importantes dans la maintenance du génome et la réplication virale. Les protéines structurales à savoir spike (S), une protéine d'enveloppe (E), la protéine membranaire (M) et la nucléocapside (N) sont situées au tiers terminal 3' du génome. En plus de ces gènes, il existe plusieurs protéines accessoires qui aident à la réplication du virus (Figure 1) (Malik et al., 2020).

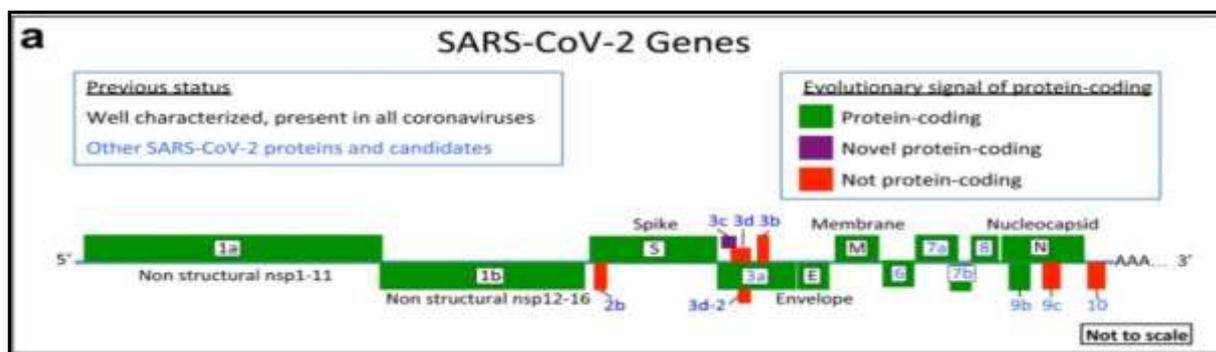


Figure 7. Les gènes du génome de SARS-CoV-2 (Jungreis et al., 2020).

1.2. Les types ORFs

- **ORF1**

Le gène ORF1 de SARS-CoV-2 couvre 21289 nucléotides, est situé entre la position 266 et 21555 du génome du virus (Tableau 2). Il est divisé en ORF1a et ORF 1ab (Yoshimoto et al., 2020).

Après l'entrée dans la cellule, la traduction de ORF1a et ORF1b de l'ARN ss (+) entier sont les premiers événements du processus infectieux (Miao et al., 2020).

L'ORF1ab contient un domaine RdRp spécifique jouant un rôle central dans la transcription et la réplication de l'ARN viral (Naqvi et al., 2020). Il exprime une polyprotéine ORF ab qui est composée de 16 protéines non structurales (NSP), alors que La polyprotéine synthétisée à partir d'ORF1a est transformée en onze protéines non structurales (NSP1-NSP11) (Miao et al., 2020 ; Yoshimoto et al.,2020).

Les protéines non structurelles ORF 1a et 1ab partagent la majorité du contenu nucléotidique du génome viral. Les polyprotéines pp1a et pp1ab sont codées par les ORF 1a et 1b (Naqvi at al., 2020).

- **ORF 2**

Le gène ORF2 couvre 3821 nucléotides, il est situé entre la position 21563 et 25384 du génome du virus (Yoshimoto et al., 2020). Ils sont nommés dans le dernier tiers du génome codent pour la glycoprotéine S de surface (ORF2), qui est clivé en S1 et S2 et est responsable de l'attachement et de l'entrée virale en se liant au récepteur ACE2 humain (Jungereis et al., 2020).

- **ORF 3**

Le gène ORF3 couvre 827 nucléotides, il est situé entre la position 26220 et 25393 du génome du virus. Le gène ORF 3a est codé pour la protéine ORF 3a (protéine accessoire) (Yoshimoto et al., 2020).

- **ORF 4**

Le gène ORF4 couvre 227 nucléotides, est situé entre la position 26245 et 26472 du génome du virus. Le gène ORF 4 est codé pour la protéine E (Yoshimoto et al., 2020).

- **ORF 5**

Le gène ORF5 couvre 668 nucléotides, il est situé entre la position 26523 et 27191 du génome du virus. Le gène ORF 5 est codé pour la protéine M (Yoshimoto et al., 2020).

- **ORF 6**

Le gène ORF6 couvre 185 nucléotides, il est situé entre la position 27202 et 27387 du génome du virus. Le gène ORF 6 est codé pour la protéine ORF 6 (Yoshimoto et al., 2020).

- **ORF 7a**

Le gène ORF7a couvre 365 nucléotides, il est situé entre la position 27394 et 27759 du génome du virus. Le gène ORF 7a est codé pour la protéine ORF 7a (Yoshimoto et al., 2020).

- **ORF 7b**

Le gène ORF 7b couvre 131 nucléotides, il est situé entre la position 27756 et 27887 du génome du virus. Le gène ORF7b est codé pour la protéine ORF 7b. Il est situé 27 nucléotides en amont du codon d'arrêt ORF8. La protéine ORF7b prédite est de 7 acides aminés plus courte que la variante normale et se termine par Thr-Phe-Val-Leu-Phe (Pereira, 2020 ; Yoshimoto et al., 2020).

- **ORF 8**

Le gène ORF8 couvre 366 nucléotides, il est situé entre la position 27 894 et 28 259 du génome du virus (Yoshimoto et al., 2020), il est suivant l'ORF7b et précédant les gènes N, respectivement, et code pour une protéine ORF8 longue de 121 acides aminés (Zinzula, 2021).

- **ORF 9**

Le gène ORF9 couvre 1259 nucléotides (nt), il est situé entre la position 28274 et 29533 du génome du virus. Le gène ORF 9 est codé pour la protéine N (Yoshimoto et al., 2020).

- **ORF 10**

Le gène ORF10 couvre 116 nucléotides, il est situé entre la position 29558 et 29674 du génome du virus, et code pour la protéine ORF10 (Yoshimoto et al., 2020).

Tableau 2. Résume la liste des gènes ORFs du SARS-CoV-2 et leurs localisation et protéines

Tableau 2. L'expression des gènes de SARS-CoV-2 (NC045512.2) (Yoshimoto et al., 2020)

Gene	Gene ID	Location	Protéine	Locus
ORF1ab	43,740,578	266-21,555	Polyprotéine ORF1ab	[BCB15089.1/BCB97900.1]
ORF1a	43,740,578	26613,483	Polyprotéine ORF 1a	[YP_009725295.1]
ORF2	43,740,568	21,563-25,384	Protéine S	[BCA87361.1]
ORF3a	43,740,569	25,393–26,220	Protéine 3a	[BCA87362.1]
ORF4	43,740,570	26,245–26,472	Protéine E	[BCA87363.1]
ORF5	43,740,571	26,523–27,191	Protéine M	[BCA87364.1]
ORF6	43,740,572	27,202–27,387	Protéine ORF6	[BCA87365.1]
ORF7a	43,740,573	27,394–27,759	Protéine ORF7a	[BCA87366.1]
ORF7b	43,740,574	27,756–27,887	Protéine ORF 7b	[BCB15096.1]
ORF8	43,740,577	27,894–28,259	Protéine ORF8	[BCA87367.1]
ORF9	43,740,575	28,274–29,533	Protéine N	[BCA87368.1]
ORF10	43,740,576	29,558–29,674	Protéine ORF 10	[BCA87369.1]

2. Les protéines non structurales

Le pp1a et pp1b est fragmentés par des protéases virales en 16 protéines non structurales (NSPs), qui sont assemblé en complexe réplique-transcriptase et présentent une activité enzymatique multiple (Gorla et al., 2020 ; Yan et al., 2020).

NSP1 à NSP11 sont codés par ORF1a et sont donc présents à la fois dans pp1a et pp1ab, alors que NSP12 à NSP16 sont codés par ORF1b et sont uniquement présent dans pp1ab (Llanes et al., 2020). Les NSPs sont impliqués tout au long des phases initiales et intermédiaires du cycle de vie viral (Gadhavé et al., 2020), parmi eux, NSP12 fonctionne comme ARN polymérase ARN-dépendante (RdRp) par en utilisant NSP7 et NSP8 comme facteurs auxiliaires, et NSP13 est identifié comme une hélicase. Le NSP5 possède une activité protéase de type chymotrypsine (Santerre et al., 2020 ; Yan et al., 2020).

2.1. Les types NSPs

- **NSP 1**

La protéine non structurale 1 est la première protéine de la polyprotéine (pp1a) de SRAS- CoV-2 Cette protéine est également connue sous le nom de protéine leader. Cette protéine se trouve également dans le coronavirus du SARS (Yoshimoto et al., 2020), et il peut jouer des rôles spécifiques dans l'interaction du virus avec la réponse immunitaire innée (Tableau 3) (Graham et al., 2008 ; Gorla et al.,2020)

Il est connu pour être un puissant inhibiteur d'expression du gène hôte, il se lie au ribosome 40S de la cellule hôte qui dégrade les ARNm de l'hôte en induisant un clivage endonucléolytique près de la région 5' non traduite et inhibe ainsi la traduction de l'hôte (tandis que l'ARNm du virus L'ARNm reste intact. La protéine leader favorise une expression de gènes viraux dans les cellules infectées et échapper de la réponse immunitaire de l'hôte (Gorla et al 2020 ; Yoshimoto et al., 2020).

- **NSP 2**

La protéine non structurale 2 est la deuxième protéine de la polyprotéine (pp1a) de SARS-CoV-2. Dans SARS-CoV-2, NSP2 s'est lié à deux hôtes protéines : prohibitin 1 et prohibitin 2 (PHB1 et PHB2) Les protéines PHB1 et PHB2 (répresseur de l'activité des récepteurs des œstrogènes) sont omniprésents exprimé et présent dans plusieurs compartiments cellulaires comme les mitochondries sont connues pour jouer un rôle dans progression du cycle cellulaire, migration cellulaire, différenciation cellulaire, l'apoptose et la biogenèse mitochondriale. La reliure des protéines NSP2 en PHB1 et PHB2 suggèrent que NSP2 joue un rôle dans la perturbation de l'environnement de la cellule hôte (Gorla et al 2020 ; Yoshimoto et al., 2020).

- **NSP 3**

Le NSP3 est la protéine protéinase de type papaine. Cette protéine a une taille de près de 200 KDa et est la plus grande protéine (Yoshimoto et al., 2020), avec un domaine de noyau protéolytique de 316 acides aminés responsable du clivage des trois premiers sites de clivage à l'extrémité N-terminale des substrats de réplicase, il a une activité qui supprime la réponse immunitaire innée antivirale de la cellule hôte, il empêche également la signalisation du facteur nucléaire Kappa B (NF-κB) de l'hôte et antagonise la réponse immunitaire innée de l'interféron de type I en bloquant la phosphorylation, la dimérisation et la translocation nucléaire ultérieure de IRF3. La mutation déstabilisante qui se produit près du domaine phosphatase du NSP3 peut suggérer un mécanisme potentiel qui différencie le SARS-CoV-2. En raison de ses diverses actions, NSP3 peut fournir de nouvelles voies pour étudier le cycle de réplication du virus dans les cellules hôtes, dans le but de développer des agents thérapeutiques pour inhiber la réplication du SARS-CoV-2 (Gorla et al., 2020).

NSP 3 et NSP 4 interagissent avec d'autres cofacteurs pour induire un réarrangement membranaire pour le mécanisme de réplication virale et la perte du complexe NSP 3- NSP 4 élimine la réplication virale (Raj, 2021).

- **NSP 4**

NSP4 glycosylé qui interagit avec NSP3 et éventuellement des protéines hôtes pour confèrent un rôle lié au réarrangement des membranes dans le SARS-CoV. De plus, l'interaction entre NSP4 et NSP3 est essentielle pour la réplication virale grâce aux réarrangements de double membrane cytoplasmique induite par le virus vésicules (Gorla et al., 2020 ; Santerre et al., 2020 ; Yoshimoto et al., 2020).

- **NSP 5**

NSP 5 est une protéase de type chymotrypsine (3CL-PRO) ou Main protéase (M pro), une enzyme clé responsable de réplication du SARS-CoV-2. Le M pro existe en homodimère et a une dyade Cys-His sur un site actif qui montre une protéase activité. (Gorla et al., 2020 ; Yoshimoto et al., 2020).

Il est traité 11 sites de clivage entre NSP 4 et 16 pendant la réplication et a également une structure à 3 domaines conservée et des résidus catalytiques (Raj, 2021).

- **NSP 6**

La protéine non structurale 6 du SARS-CoV-2 est une protéine membranaire d'environ 34 kDa avec six hélices transmembranaires et un C-terminal il s'associe au NSP3 et NSP4 lors de l'assemblage de la réplication du coronavirus complexe pour former des DMV et se lie avec le récepteur sigma du réticulum endoplasmique de l'hôte et déclenche l'induction d'autophagosomes. Généralement l'activation du récepteur sigma régule la réponse au stress du réticulum endoplasmique. NSP3, NSP4 et NSP6 ont tous ensemble la capacité à induire des vésicules à double membrane dans lesquelles NSP3 et NSP4 sont capables de coupler des membranes tout en NSP6 a une capacité de prolifération membranaire (Gorla et al., 2020 ; Santerre et al., 2020).

- **NSP 7 et NSP8**

NSP7 se lie avec NSP8 (8 sous-unités de chaque) pour former un hexadécamère connu sous le nom de complexe primase qui interagit avec l'ARN polymérase ARN-dépendante et NSP7 et NSP8 aident à la processivité de RdRp et forment ensemble l'holoenzyme polymérase principale. Elles forment un complexe hétéro-oligomérique (NSP12-NSP7 / NSP8) qui peuvent participer à la réplication du SARS-CoV-2 (Gorla et al., 2020 ; Llanes et al., 2020).

- **NSP 9**

NSP9 en complexe avec NSP8 et elle agit comme une protéine de liaison à l'ARN simple brin qui impliquée dans la réplication virale et la virulence (Raj, 2021).

NSP 9 liée l'ARN par l'oligosaccharide / pli en forme de pli oligonucléotidique unique à la classe des β CoV jouer un rôle similaire dans le maintien de l'intégrité fonctionnelle de SARS-CoV-2. Les chercheurs ont déterminé la structure de NSP9 et identifié un site de liaison peptidique qui pourrait inciter davantage recherche pour comprendre son rôle dans COVID-19 (Gorla et al., 2020 ; Raj, 2021).

- **NSP 10**

Le complexe NSP10 - NSP16 est essentiel pour coiffer les transcriptions d'ARNm viral pour une traduction efficace et pour échapper à la surveillance immunitaire (Raj; 2021).

NSP10 est un cofacteur essentiel qui active la 2' O-ribose méthyltransférase (NSP16) et il interagit avec NSP14 dans le SARS coronavirus, et cette interaction stimule l'activité de NSP14 et joue un rôle important dans la méthylation des ARNm guanosine cap pour favoriser la transcription, l'épissage, la polyadénylation et l'exportation nucléaire de l'ARNm viral. Le complexe hétéro-oligomérique (NSP10-NSP14-NSP16) peut être une cible pour le développement d'antiviraux contre les coronavirus pathogènes (Gorla et al., 2020 ; Yoshimoto et al., 2020).

- **NSP 11**

NSP11 est un peptide le plus court seulement 13 acides aminés de long, les neuf premiers acides aminés sont identiques aux neuf premiers de NSP12. La protéine NSP11 est le produit de clivage de la pp1a polyprotéine par la protéase 3CLpro /Mpro à la jonction NSP10 /11 (Gadhave et al., 2020 ; Yoshimoto et al., 2020).

- **NSP12**

NSP12 est une RdRp très essentielle de la machinerie de réplication virale et interagit avec ses cofacteurs (NSP7 et NSP8) qui ensemble stimulent la réplication génomique, donc NSP12 est incapable de remplir sa fonction en tant qu'entité unique mais doit exister dans un complexe tripartite avec NSP7 et NSP8 pour activer la réplication de longs ARN. Le complexe NSP7-NSP8 améliore la liaison de NSP12 à l'ARN en plus de son activité enzymatique (Olotu et al., 2020).

- **NSP13**

NSP13 est une sous-unité d'hélicase hautement conservée qui est nécessaire pour une réplication efficace du génome viral, il a également un N-terminal riche en cystéine domaine de liaison au zinc (ZBD) qui semble moduler l'activité de l'hélicase (Llanes et al., 2020). En plus de son activité hélicase, l'NSP13 du SARS-CoV est également connu pour posséder une activité 5'-triphosphatase, qui est responsable de l'introduction du capuchon 5'-terminal du ARNm viral (Yoshimoto et al., 2020).

- **NSP14**

Nsp14 est une protéine bifonctionnelle qui possède deux activités : une activité N7-guanine méthyltransférase (N7-MTase), impliquée dans la méthylation de la coiffe, et une activité exonucléase 3'-5, permettant l'excision de nucléotides (Llanes et al., 2020 ; Corteggiani et al., 2021).

Elle forme un dodécamère pour améliorer leurs activités enzymatiques et joue un rôle clé dans la synthèse de l'ARN et la fidélité de réplication. La partie N-terminale de NSP14 interagit avec l'ARN hélicase ATP-dépendante, la région C-terminale qui améliore la réplication (Gorla et al., 2020).

- **NSP15**

NSP15 du coronavirus du SARS a été caractérisé biochimiquement comme endoribonucléase qui clive l'ARN au niveau des uridylates en position 3' pour former un produit 2'-3'phosphodiester cyclique. La protéine NSP15 cible et dégrade spécifiquement les séquences virales de polyuridine pour empêcher le système immunitaire de l'hôte. Elle utilise le manganèse comme cofacteur pour favoriser l'activité endoribonucléase, cette activité de NSP15 est améliorée par NSP8 et le complexe NSP7/NSP8 (Gorla et al., 2020 ; Yoshimoto et al., 2020).

- **NSP16**

NSP16 joue un rôle important dans la méthylation de coiffe des ARNm viraux qui est nécessaire pour éviter le système immunitaire. L'activité enzymatique de NSP16 est augmentée par l'interaction avec NSP10 et est essentielle pour la réplication et transcription virale dans les cellules hôtes, elle méthyle le groupe 2'-hydroxy de l'adénine en utilisant S-adénosylméthionine comme source de méthyle (Gorla et al., 2020 ; Yoshimoto et al., 2020).

Le Tableau 3 résume la liste des protéines non structurales (NSP) du SARS-CoV-2 et leurs fonctions moléculaires.

Tableau 3. Les protéines non structurales et leur fonction moléculaire (Naqvi et al., 2020 ; Yoshimoto et al., 2020).

Protéines non structurales	Accession	Acides Aminés	Description	Fonctions proposées
NSP1	YP_009725297.1	180	NSP1 est le N terminal produit de la réplicase virale.	Traduction de la protéine hôte leader inhibiteur. Médiation de l'ARN Réplication et traitement. Impliqué dans la dégradation de l'ARNm.
NSP2	YP_009725298.1	638	NSP2 est un produit de réplicase essentiel pour relecture la réplication virale.	Modulation de la survie des cellules hôtes voie de signalisation en interagissant avec l'hôte PHB et PHB2
NSP3	YP_009725299.1	1945	NSP3 est un papaïne Protéinase contient plusieurs domaines.	Libérer les NSP 1, 2, 3(Papaïne comme protéinase).
NSP4	YP_009725300.1	500	Contient le domaine transmembranaire 2 (TM2).	Réarrangement de la membrane.
NSP5	YP_009725301.1	306	Protéinase de type 3C	Impliqué dans le traitement de la polyprotéine virale pendant la réplication.
NSP6	YP_009725302.1	290	Domaine transmembranaire putatif	Joue un rôle dans l'induction initiale des autophagosomes du RE de l'hôte.
NSP7	YP_009725303.1	83	ARN polymérase ARN dépendante.	Forme un complexe avec NSP8 et NSP12 pour produire l'activité ARN polymérase de NSP8.
NSP8	YP_009725304.1	198	ARN polymérase multimérique réplicase.	Il forme un super complexe hexadécamerique avec NSP7. Stimule NSP12.
NSP9	YP_009725305.1	113	Liaison ARN viral protéine.	Participe à la réplication virale.
NSP10	YP_009725306.1	139	Facteur de croissance contient deux liants zinc motifs	Joue un rôle essentiel dans la coiffe des ARNm viraux.
NSP11	YP_009725312.1	13	-	-
NSP12	YP_009725307.1	932	Dépendant de l'ARN ARN polymérase (Pol / RdRp)	Responsable de la réplication et de la transcription de l'ARN viral du génome.
NSP13	YP_009725308.1	601	Hélicase	Un domaine principal d'hélicase qui se lie l'ATP. Le domaine de liaison au zinc est impliqué dans la réplication et transcription.
NSP14	YP_009725309.1	527	Relecture du domaine Exoribonucléase (ExoN / nsp14)	Activité exoribonucléase agissant dans une direction 3' à 5' et N7-guanine-méthyltransférase.
NSP15	YP_009725310.1	346	EndoRNase	Dégrader l'ARN en (endoRNase / endoribonucléase) échapper à la défense de l'hôte.
NSP16	YP_009725311.1	298	2'-O-ribose méthyltransférase	Intervient dans la méthylation.

3. Les protéines accessoires

En plus de ces éléments génomiques (ORF1a et ORF1ab) partagés par d'autres coronavirus, le génome du SARS-CoV contient également huit ORF codant pour protéines accessoires, à savoir les ORF 3a, 3b, 6, 7a, 7b, 8a, 8b et 9b (Liu et al., 2014).

Au niveau des acides aminés, le SARS-CoV-2 est assez similaire à celui de SARS-CoV, mais il existe des différences. Par exemple, la protéine 8a est présent dans le SARS-CoV et absent dans SARS-CoV-2, la protéine 8b est 84 acides aminés dans le SARS-CoV, mais plus longtemps en SARS-CoV-2, avec 121 acides aminés, la protéine 3b représente 154 acides aminés dans le SARS-CoV, mais plus court en SARS-CoV-2 avec seulement 22 acides aminés (Wu et al., 2020).

Bien que des preuves récentes montrent que les gènes accessoires du SARS-CoV sont exprimés dans l'hôte pendant l'infection, leurs fonctions restent quelque peu obscures. Il existe une gamme de fonctions proposées pour ces protéines accessoires, y compris la modulation de la pathogénicité virale et de la réplication, ainsi que d'agir comme des inducteurs de mort cellulaire et des antagonistes de l'interféron (IFN). Actuellement, les protéines accessoires de SARS-CoV-2 ne sont souvent pas indispensables à la réplication ou à la structure virale, mais jouent un rôle dans la pathogénicité virale ou se propagent par moduler les voies de signalisation de l'interféron hôte (McBride et al., 2012 ; Michel et al., 2020).

4. Les protéines structurales

4.1. Protéine M

Les protéines M sont des protéines structurales longues de 222 acides aminés qui fonctionnent de concert avec les protéines E, N et S, et joue un rôle majeur dans l'encapsidation de l'ARN (Naqvi et al., 2020). C'est la protéine structurelle la plus abondante dans le virion et on pense qu'elle donne le virion sa forme. Des études récentes suggèrent que la protéine M existe sous forme de dimère dans le virion, et peut adopter deux conformations différentes lui permettant de favoriser également la courbure de la membrane comme se lie à la nucléocapside. Elle comprend un court N-terminal glycosylé ectodomaine avec trois domaines transmembranaires (TM) et un long domaine CT C-termina (Figure 8). Elle est également considérée comme l'organisateur central de

l'assemblage des coronavirus, interagissant avec tous les autres protéines structurales corona virales (Pathak, 2020).

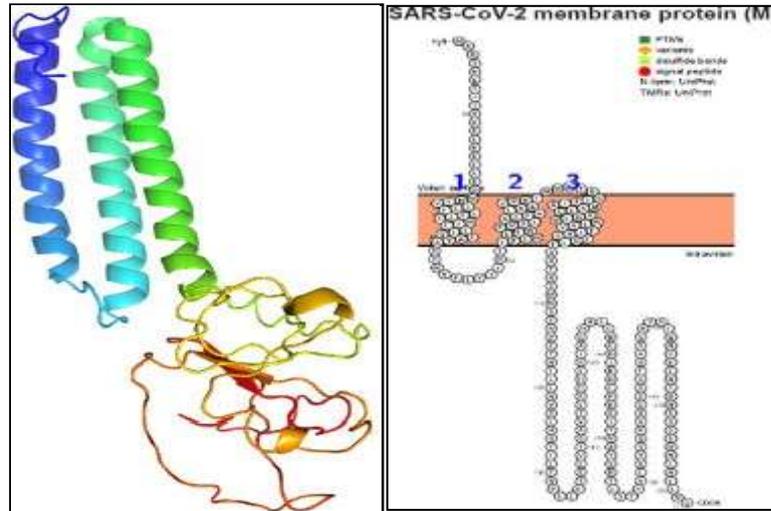


Figure 8. Structure de la protéine M (Thomas, 2020).

4.2. Protéine E

La protéine E est une protéine membranaire courte et intégrale qui contient 76 à 109 résidus d'acides aminés (Figure 9), d'une taille comprise entre 8,4 et 12 kDa, et se compose de 35 hélices α et 40 boucles. La protéine a une extrémité N-terminale hydrophile courte constituée de 7 à 12 résidus d'acides aminés, suivie d'un grand domaine transmembranaire hydrophobe de 25 résidus d'acides aminés et d'un long domaine C-terminal hydrophile (Wabalo et al., 2020).

On la trouve en petites quantités dans le virion. Les E-protéines des coronavirus sont très divergents mais ont une architecture commune. Elle facilite l'assemblage et la libération du virus mais a également d'autres fonctions telles que l'activité des canaux ioniques dans la protéine E du SARS-CoV est non requis pour la réplication virale mais requis pour la pathogénèse (Pathak, 2020).

Elle est considérée comme une cible médicamenteuse potentielle, et jouent un rôle important dans la morphogénèse virale et assemblage. La protéine E est connue pour agir comme des viroporines qui s'assemblent en hôte membrane formant des pores protéines-lipides impliqués dans le transport ionique (Naqvi et al., 2020).

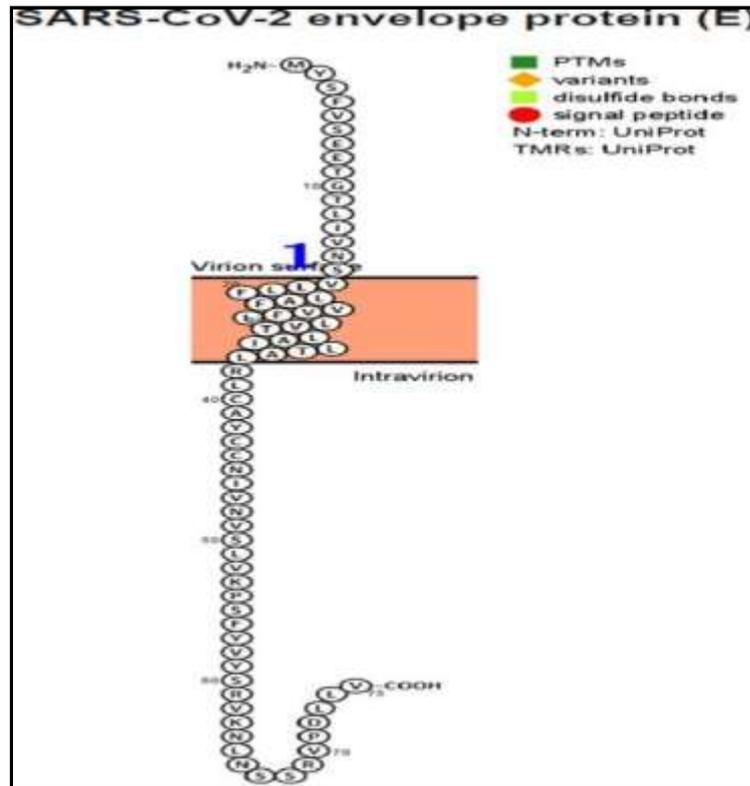


Figure 9. Structure de la protéine E (Thomas, 2020).

4.3. La protéine N

Les protéines N du bêta-coronavirus partagent une structure de domaine globale commune (Ye et al., 2020).

La protéine nucléocapside du SARS-CoV-2 est une protéine de liaison à l'ARN multivalente de 46 kDa, qui contient un domaine N-terminal replié qui participe au domaine de liaison à l'ARN (NTD) précédé d'une région désordonnée N-terminale de 44 acides aminés (NIDR) et suivi d'un lieu de 73 acides aminés riche en sérine / arginine. Le lieu flexible est adjacent au domaine de dimérisation replié (CTD) suivi de la région de queue désordonnée 52 amino C-terminale (CIDR) riche en lysine et en glutamine (Figure 10) (Perdikari et al., 2020).

Elle est impliquée dans des processus liés au génome viral, au cycle de réplication virale et à la réponse cellulaire des cellules hôtes aux infections virales. La protéine N est également fortement phosphorylée et suggérée pour conduire à des changements structurels améliorant l'affinité pour l'ARN viral (Wabalo et al., 2020).

atomique. La région S1 contient le NTD, le CTD et contient le domaine RBD (qui assure la reconnaissance du récepteur cellulaire ACE2 par le virus. Elle contient également des sites antigéniques qui sont exposés à la surface du virus et accessibles au système immunitaire, constituant des antigènes pouvant être reconnus par les anticorps produits par les hôtes infectés. Cependant, la séquence génomique codant pour ces séquences peptidiques potentiellement antigéniques présente une grande variabilité entre les espèces virales qui résulte de la sélection de mutations génétiques permettant aux virus d'échapper à la réponse immunitaire de l'hôte. Les résidus du RBD impliqués directement dans la reconnaissance d'ACE2 (Perdikari et al., 2020).

La sous-unité S2, composée successivement d'un domaine FP, HR1, HR2, TM et d'une fusion de domaine cytoplasmique (CT), est responsable de la fusion et de l'entrée virale. FP est un court segment de 15 à 20 acides aminés conservés de la famille virale, composé principalement de résidus hydrophobes, tels que la glycine (G) ou l'alanine (A), qui s'ancrent à la membrane cible lorsque la protéine S adopte la conformation de la pré-épine. Des recherches antérieures ont montré que la FP joue un rôle essentiel dans la médiation de la fusion membranaire en perturbant et en reliant les bicouches lipidiques de la membrane de la cellule hôte. Les HR1 et HR2 sont composés d'un heptapeptide répétitif : HPPHCPC, où H est un résidu hydrophobe ou traditionnellement volumineux, P est un résidu polaire ou hydrophile et c'est un autre résidu chargé. HR1 et HR2 forment le faisceau à six hélices qui est essentiel pour la fusion virale et la fonction d'entrée de la sous-unité S2. HR1 est situé à l'extrémité C-terminale d'un FP hydrophobe, et HR2 est situé à l'extrémité N-terminale du domaine TM. Le domaine TM en aval ancre la protéine S à la membrane virale et la sous-unité S2 se termine par une queue CT (Huang et al., 2020).

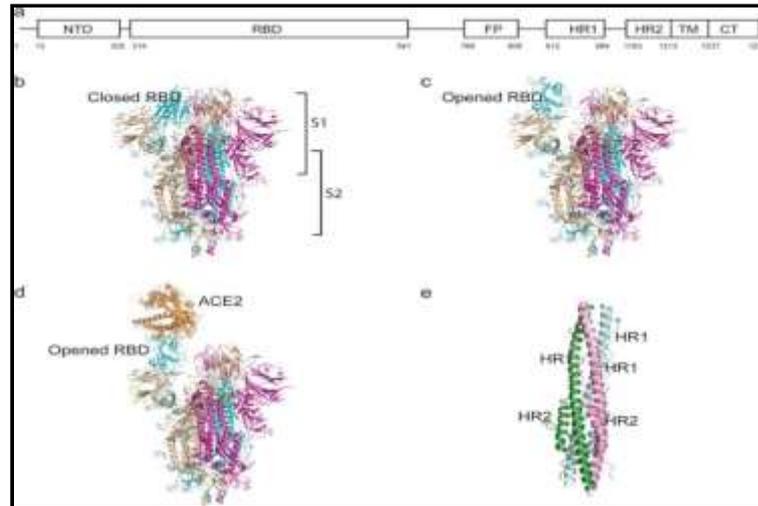


Figure 11. Structure de la protéine S (Huang et al., 2020).

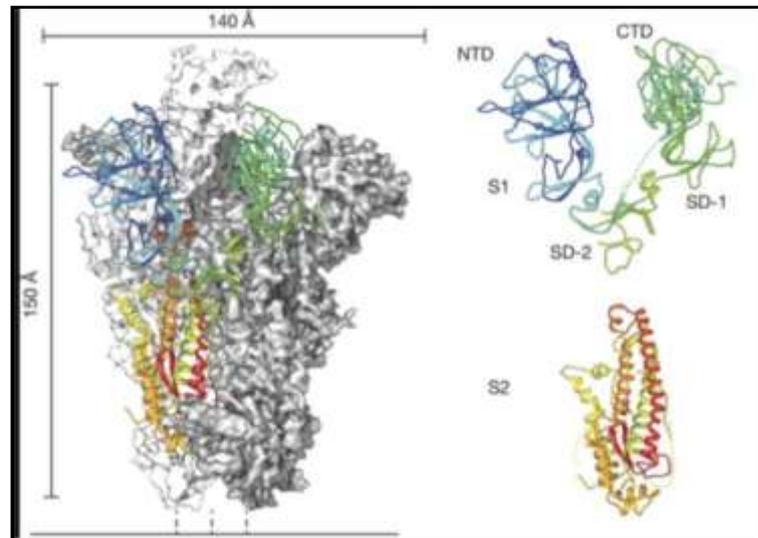


Figure 12. Structure 3D de la protéine S (PDB, 2021).

5. Phylogénie du SARS-CoV-2

5.1. Phylogénie du génome complet

Dans une analyse de réseau phylogénétique de 160 génomes complets du SARS-CoV-2, trois variantes centrales distinguées sont trouvées par des changements d'acides aminés, qui les nommés A, B et C, A étant le type ancestral selon l'exogroupe chauve-souris coronavirus. Les types A et C se trouvent dans des proportions importantes en dehors de

l'Asie de l'Est, c'est-à-dire chez les Européens et les Américains. En revanche, le type B est le type le plus courant en Asie de l'Est, et son génome ancestral ne semble pas s'être propagé en dehors de l'Asie de l'Est sans d'abord muter en types B dérivés, indiquant des effets fondateurs ou une résistance immunologique ou environnementale contre ce type en dehors de l'Asie. Le réseau retrace fidèlement les voies d'infection pour les cas documentés de la maladie à coronavirus 2019 (COVID-19), indiquant que les réseaux phylogénétiques peuvent également être utilisés avec succès pour aider à retracer les sources d'infection COVID-19 non documentées, qui peuvent ensuite être mises en quarantaine pour empêcher la propagation récurrente de la maladie à l'échelle mondiale (Forster et al., 2020).

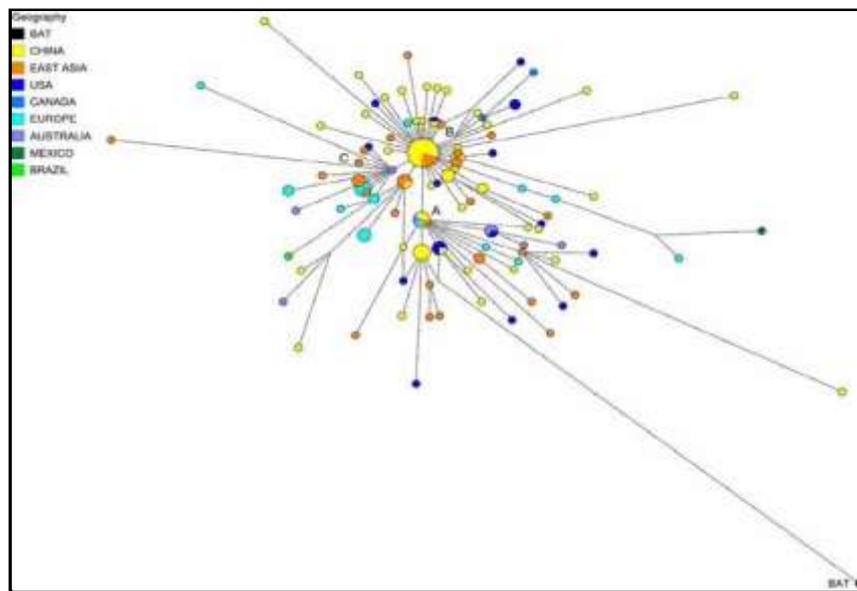


Figure 13. Réseau phylogénétique de 160 génomes du SARS-CoV-2 (Forster et al., 2020).

Pour étudier l'histoire évolutive de la récente épidémie du SARS-CoV-2 en Chine, un total de 70 génomes de souches virales de Chine et d'ailleurs avec des dates d'échantillonnage entre le 24 décembre 2019 et le 3 février 2020 ont été analysés. Pour explorer l'hôte animal intermédiaire potentiel du virus SARS-CoV-2, les ensembles de données de virome de pangolins et d'isolats représentatifs de coronavirus liés au SARS provenant de chauves-souris ont été réanalysés, en accordant une attention particulière au gène de la glycoprotéine de pointe. Des analyses phylogénétiques ont été effectuées, de réseau fractionné, de réseau de transmission, de cartographie de vraisemblance et comparative des génomes. Sur la base d'une analyse phylogénétique bayésienne à l'échelle du temps à l'aide de la méthode de datation par pointe, le temps est estimé jusqu'à l'ancêtre

commun le plus récent et le taux d'évolution du SARS-CoV-2, qui variait du 22 au 24 novembre 2019 et de $1,19$ à $1,31 \times 10^{-3}$ remplacements par site et par an, respectivement. Les résultats ont également révélé que le virus BetaCoV/bat/Yunnan/RaTG13/2013 était plus similaire au virus SARS-CoV-2 que le coronavirus obtenu à partir des deux échantillons de pangolin (SRR10168377 et SRR10168378). Une insertion est identifiée de peptide unique (PRRA) dans le virus SARS-CoV-2 humain, qui pourrait être impliquée dans le clivage protéolytique de la protéine de pointe par les protéases cellulaires, et pourrait donc avoir un impact sur la gamme d'hôtes et la transmissibilité. Fait intéressant, le coronavirus porté par les pangolins n'avait pas le motif RRAR. Par conséquent, en conclusion le virus humain SARS-CoV-2, qui est responsable de la récente épidémie de COVID-19, ne provenait pas directement des pangolins (Li et al., 2020).

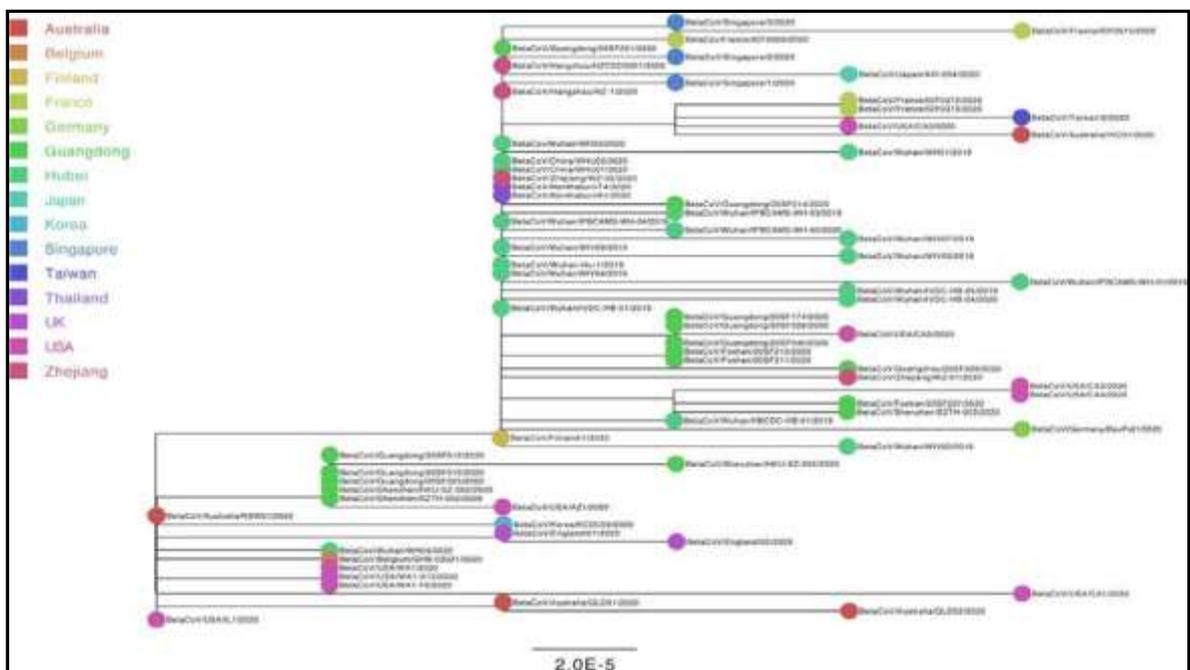


Figure 14 Estimation de la phylogénie du maximum de vraisemblance du coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2). Les couleurs indiquent les différents emplacements d'échantillonnage. L'arbre est enraciné au milieu. Les résultats des analyses phylogénétiques de vraisemblance maximale de « dataset-70 » sont affichés (Li et al., 2020).

5.2. Phylogénie du génome partiel

L'objectif de l'étude de Tabibzadeh et al. (2020) était d'étudier et de suivre le SARS-CoV-2 chez les patients atteints de la maladie à coronavirus iranien 2019 (COVID-19) à l'aide de méthodes moléculaires et phylogénétiques.

Sept cas confirmés de patients COVID-19 ont fait l'objet d'une évaluation phylogénétique du SARS-CoV-2 en Iran. Les gènes nsp-2, nsp-12 et S ont été amplifiés à l'aide de la RT-PCR en une étape et séquencés à l'aide de la méthode de séquençage de Sanger. Des logiciels de bio-informatique populaires ont été utilisés pour l'alignement et l'analyse des séquences ainsi que pour la construction phylogénétique.

L'âge moyen des patients de la présente étude était de $60,42 \pm 9,94$ ans et 57,1 % (4/7) étaient des hommes. Les résultats ont indiqué une grande similitude entre les souches iraniennes et chinoises. Il n'y a pas de polymorphismes particuliers dans les régions évaluées des trois gènes. Les arbres phylogénétiques par la méthode d'association des voisins et du maximum de vraisemblance des gènes nsp-2, nsp-12 et S ont montré qu'il n'y avait aucune différence entre les isolats iraniens et ceux d'autres pays.

En tant qu'étude phylogénétique préliminaire sur les isolats iraniens du SARS-CoV-2, ces isolats sont étroitement liés aux séquences chinoises et de référence. De plus, aucune différence sensible n'a été observée entre les isolats iraniens et ceux d'autres pays. D'autres investigations sont recommandées en utilisant des méthodes plus complètes et des échantillons de plus grande taille (Tabibzadeh et al., 2020).

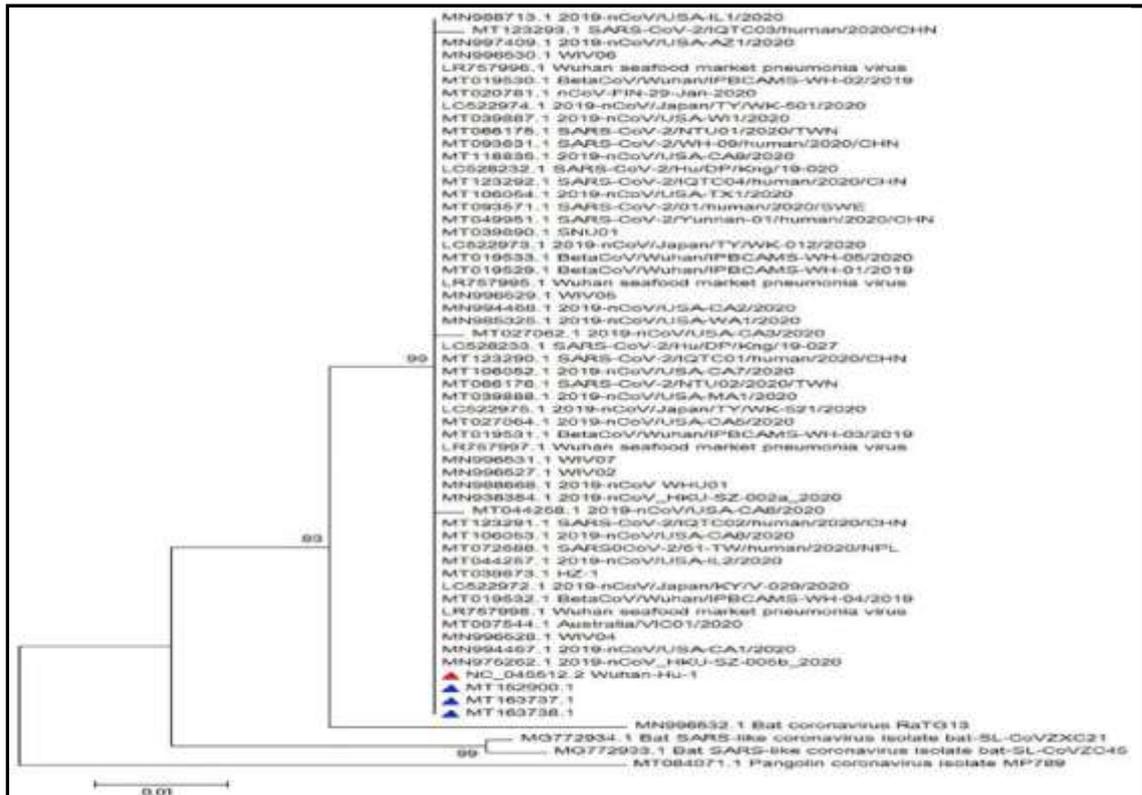


Figure 15 Arbre phylogénétique pour 322 nucléotides du gène nsp-2 du SARS-CoV-2 en utilisant la méthode de jointure voisine et 1000 bootstrap, le triangle rouge montre la séquence de référence pour le SARS-CoV-2 et le triangle bleu indique Iranien isole. (Pour l'interprétation des références à la couleur dans cette légende de la figure, le lecteur est renvoyé à la version Web de cet article) (Tabibzadeh et al., 2020)

Partie II :
Recherche
Expérimentale

Matériel
et
Méthodes

1. Matériel et Méthodes

1.1. Présentation des bases de données

1.1.1. Présentation de NCBI

Le Centre National d'Information sur la Biotechnologie (NCBI) c'est la base de données des séquences de référence (RefSeq) (<http://www.ncbi.nlm.nih.gov/RefSeq/>). Elle fournit une collection non redondante de séquences représentant les données génomiques, transcrits et protéiques. Bien que l'objectif est de fournir un ensemble de données complètes représentant les informations de séquence complètes pour toutes les espèces, la base de données comprend de manière pragmatique données de séquence actuellement accessibles au public dans les bases de données d'archives. La base de données intègre les données de plus de 2400 organismes et comprend plus d'un million de protéines représentant une importante taxonomie couvrant les procaryotes, eucaryotes et virus. Les séquences nucléotidiques et protéiques sont explicitement liées, les séquences sont liées à d'autres ressources y compris la visionneuse de cartes NCBI et Gène. Les séquences sont annotées pour inclure le codage régions, domaines conservés, variation, références, noms, références croisées de bases de données et autres fonctionnalités en utilisant une approche combinée de collaboration et d'autres contributions de la communauté scientifique, automatisées annotation, propagation depuis GenBank et curation par le personnel du NCBI (Pruitt et al., 2005).

1.1.2. Présentation GISAID

L'initiative mondiale sur le partage des données sur la grippe aviaire (GISAID) a été fondée en 2006, et depuis 2010, est hébergée par le ministère fédéral allemand de l'Alimentation, de l'Agriculture et de la Protection des consommateurs. GISAID est également devenu un référentiel de coronavirus depuis décembre 2019. Au 4 mars 2020, date limite de notre analyse phylogénétique, la base de données GISAID (<https://www.gisaid.org/>) avait compilé 254 génomes de coronavirus, isolés de 244 humains (Forster et al., 2020).

1.2. Collecte des séquences génomiques à partir des bases de données

Les séquences du génome SARS-CoV-2 ont été rassemblées à partir de NCBI (www.megasoftware.net) et GISAID (<https://www.gisaid.org/>) sous format FASTA. Un

total de 447 240 séquences de génome complet et partiel du SARS-CoV-2 a été trouvé, parmi elles 42 séquences (Tableau 4) ont été sélectionnées dans la présente étude.

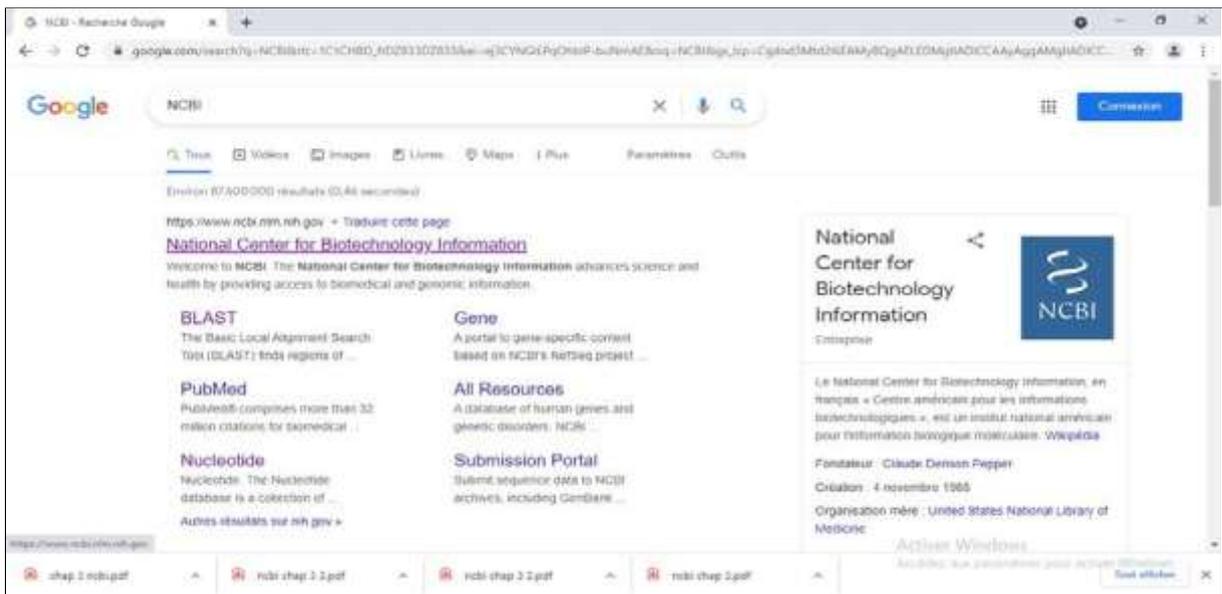
Tableau 4. Les séquences génomiques à partir des bases de données.

N°	Continent	Pays	NA	Année	Références
1	Afrique	Algeria	ESP ISL418241	02-03-2020	GSAID.
2		Algria	ESP ISL766875	17-06-2020	GSAID.
3		Algérie	EPI ISL 766864	15-07-2020	GSAID.
4		Tunisia	MT499220	2020	Fares.w et Triki.H.
5		Togo	MZ215671	2021	Halatoka et al.
6		South Africa	MT324062	2020	Allam et al.
7		Ghana	MW739939	2021	Morang'a et al.
8		Egypt	MW186669	2020	Roshdy et al.
9		Benin	MT653071	2020	Sander et al.
10	Amérique	Mexico	MW884227	2021	Vega Megana et al.
11		USA/Washington	MW811435	2020	Harcourt et al.
12		Brazil	MZ264787	2020	Cunha et al.
13		Urguay	MT466071	2020	Panzera et al.
14		Jamaica	MT507794	2020	Queen et al.
15		Colombia	MT470219	2020	Lopez et al.
16	Europe	Italy	MT066156	2020	Capobianchi et al.
17		Serbia	MT459979	2020	Vidanovic et al.
18		Spain	MW056034	2020	Cilla et al.
19		France	MT470173	2020	Deprost et al.
20		Jermany	MT270103	2020	Boehmer et al.
21		Russia	MZ047312	2020	Bocharova et al.
22		Netherlands	MT396266	2020	Oreshkova et al.
23		Poland	MT511072	2020	Sanak et al.
24		Sweden	MT093571	2020	Bengner et al.
25		Denmark	MZ049598	2020	Ramirez et al.
26	Asie	China	NC045512	2020	Wu et al.
27		India	MZ227545	2020	Kumar et al.
28		Iraq	MW633517	2021	Yasin et al.
29		Japan	LC549340	2020	Kumagai et al.
30		Jordan	MZ266636	2021	Alanagreh, L.
31		Turkey	MW306668	2020	Hacioglu et al.
32		Iran	MT320891	2020	Zeinali et al.
33		South Korea	MT039890	2020	Park et al.
34		Kazakhstan	MT428552	2020	Shevtsov et al.
35		Saudi Arabia	MZ149269	2021	Alhamlan et al.
36		Nablus	MZ207866	2021	Ereqat et al.
37		Thaïlande	MT447159	2020	Rodpan et al.
38		Pakistan	MT240479	2020	Javed et al.

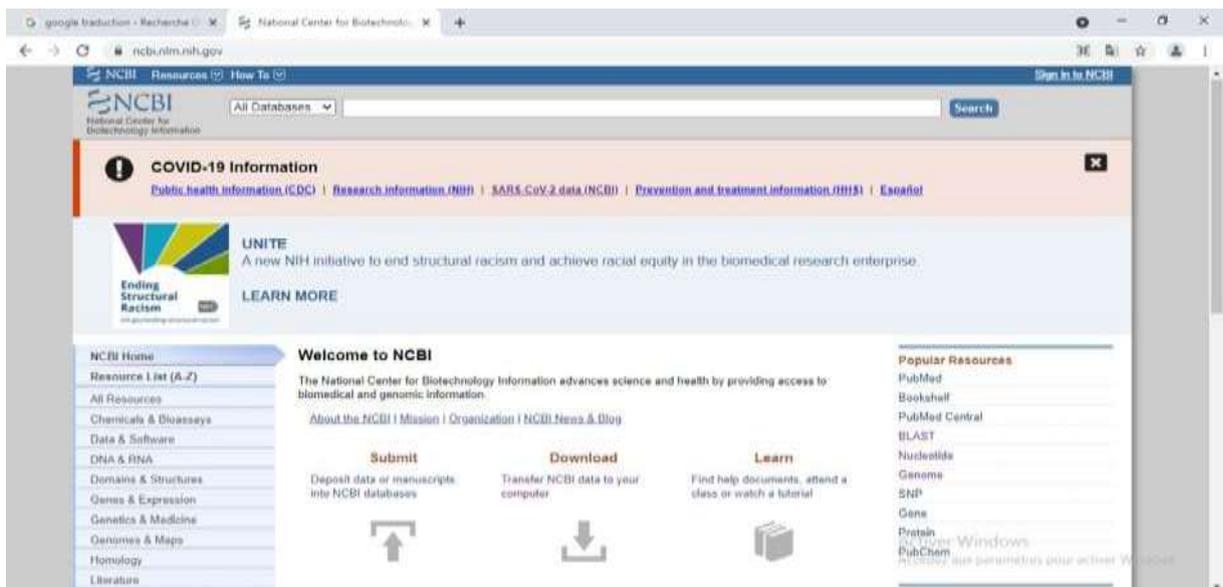
39		Viet Nam	MT192772	2020	Cao et al.
40		Taiwan	MT517436	2020	Tsao et al.
41	Océanie	Australia	MT451718	2020	Seemann et al.
42		Guam	MT459986	2020	Voegtly et al.

1.3. Etapes de la recherche sur NCBI

Les illustrations suivantes résument les étapes de la recherche de séquences génomiques sur NCBI :

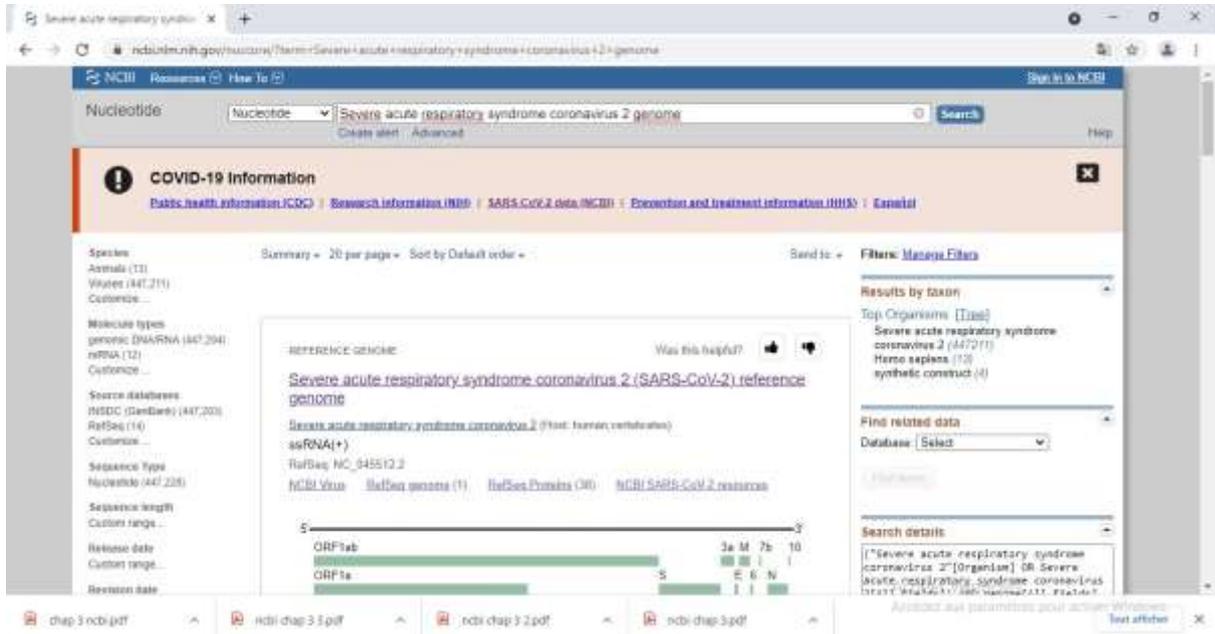


Etape 1. Ouverture de la page du navigateur Google (<https://www.ncbi.nlm.nih.gov>).



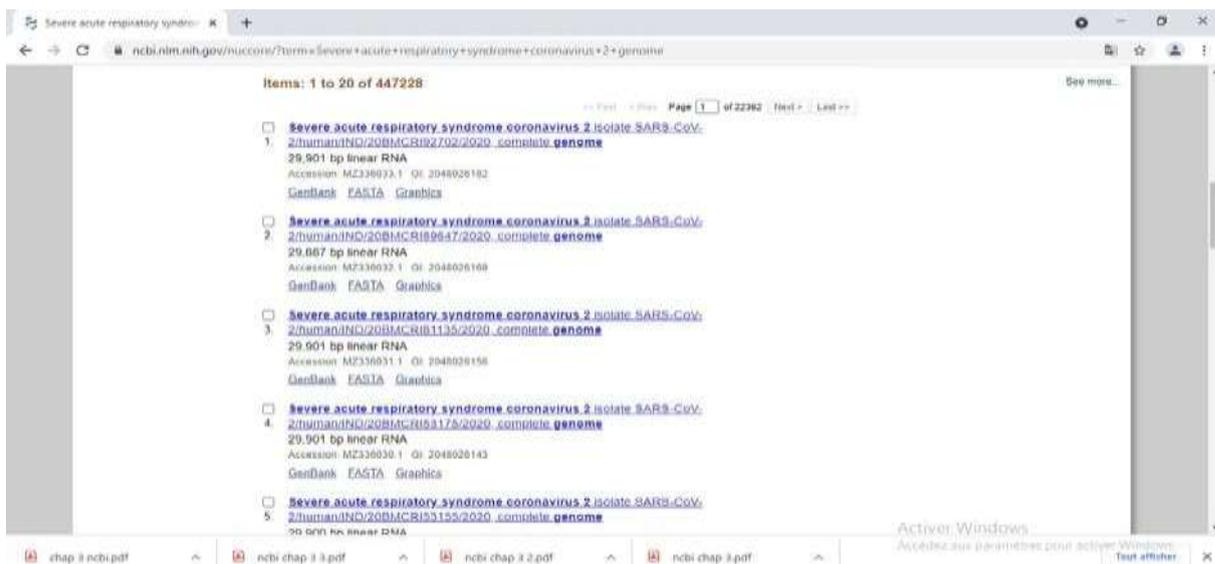
Etape 2. Page d'accueil d'NCBI

- La recherche de séquences génomique s'effectue en tapant les mots clés du génome complet du virus SARS-CoV-2 dans le champ de recherche et en cliquant sur Rechercher.

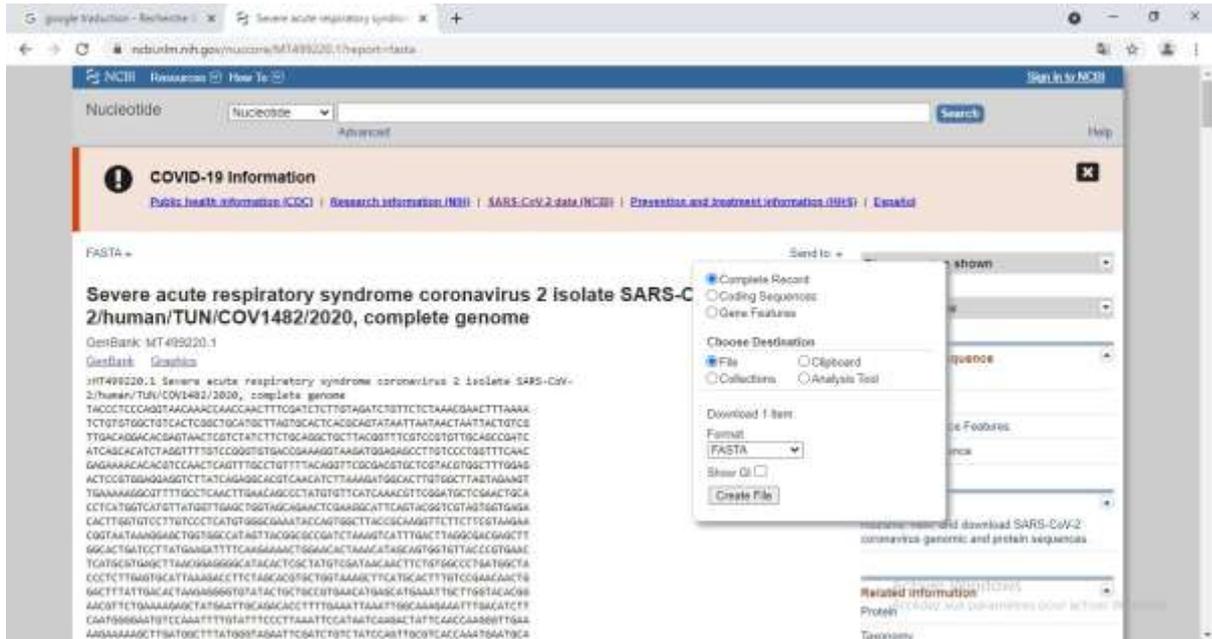


Etape 3. Saisie des mots-clés.

- Après saisie des mots clés, toutes les séquences génomiques de SARS-CoV-2 sont chargées dans chaque page contenant 20 séquences. Le totale des séquences chargées dans NCBI est 447 240 séquences de génome complet et partiel du SARS-CoV-2.



Etape 4. Les résultats de recherche du génome du SARS-CoV-2 sur NCBI.



Etape 7. Téléchargement des séquences de génome complet de SARS -CoV-2 sous format FASTA.

1.4. Les étapes de construction de l'arbre phylogénétique

1.4.1. Présentation de MEGA X

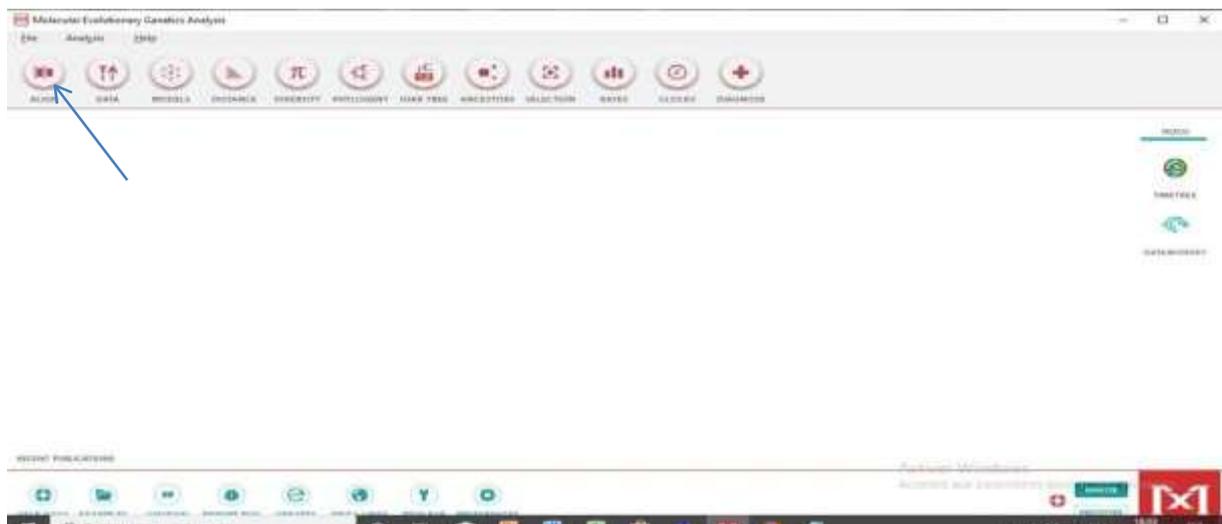
Le Logiciel Molecular Evolutionary Genetics Analysis (MEGA) met en œuvre de nombreuses méthodes et outils analytiques pour la phylogénomique et la phylomédecine. Il y'a une transformation de méga pour permettre une utilisation multiplateforme sur les systèmes d'exploitation Microsoft Windows et Linux. MEGA X ne nécessite pas de logiciel de virtualisation ou d'émulation et offre une expérience utilisateur uniforme sur toutes les plateformes. MEGA X a également été mis à niveau pour utiliser plusieurs cœurs de calcul pour de nombreuses analyses évolutives moléculaires. MEGA X est disponible en deux interfaces (graphique et ligne de commande) et peut être téléchargé gratuitement sur www.megasoftware.net, sous Microsoft Windows MEGA X est destiné à être utilisé sur les systèmes 64 bits ayant la version 10 et les versions ultérieures, le méga X multiplateforme profitera à de nombreux utilisateurs existants et permettra à de nombreux autres d'envisager d'utiliser le MEGA X dans leurs efforts de recherche et d'enseignement. La cohérence de l'expérience utilisateur entre les plateformes devrait être particulièrement utile pour faciliter les analyses collaboratives au sein et entre les groupes de recherche, ainsi que pour enseigner à un groupe diversifié d'étudiants (Figure 16) (Kamar et al., 2018).



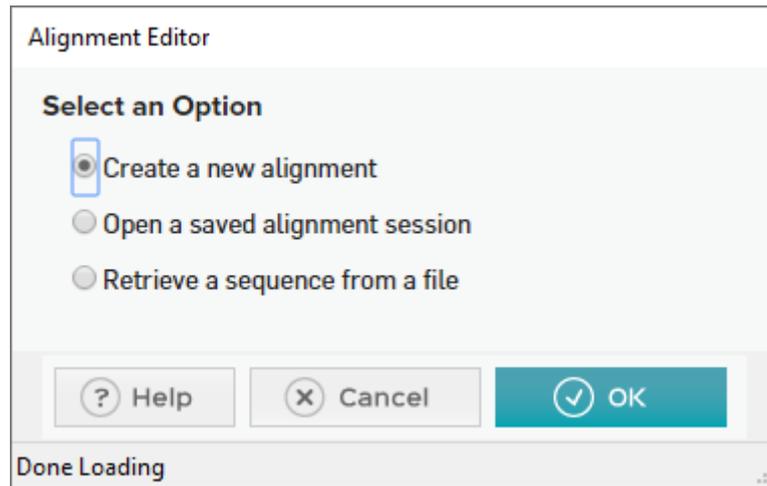
Figure 16. La fenêtre principale de MEGA X.

1.4.2. Alignement des séquences par muscle

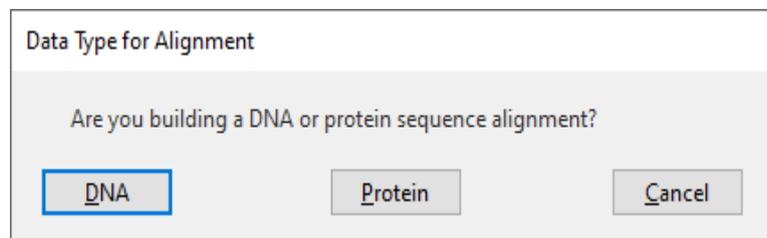
- Lancez l'explorateur d'alignement en sélectionnant Aligner / Edit /Build Alignment dans la barre de lancement de la fenêtre principale de MEGA.



Etape 8. Ouvrir le programme de MEGA X, cliquer sur Align et sélection de Edit /Build Alignment.

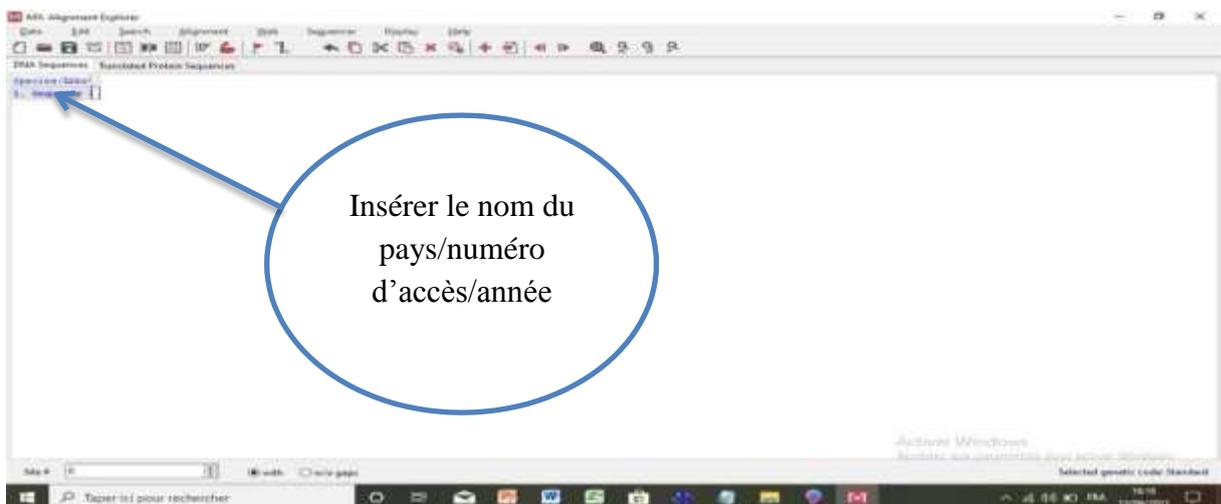


Etape 9. Choisir create a new alignment et cliquer sur OK.



Etape 10. Cliquer sur DNA.

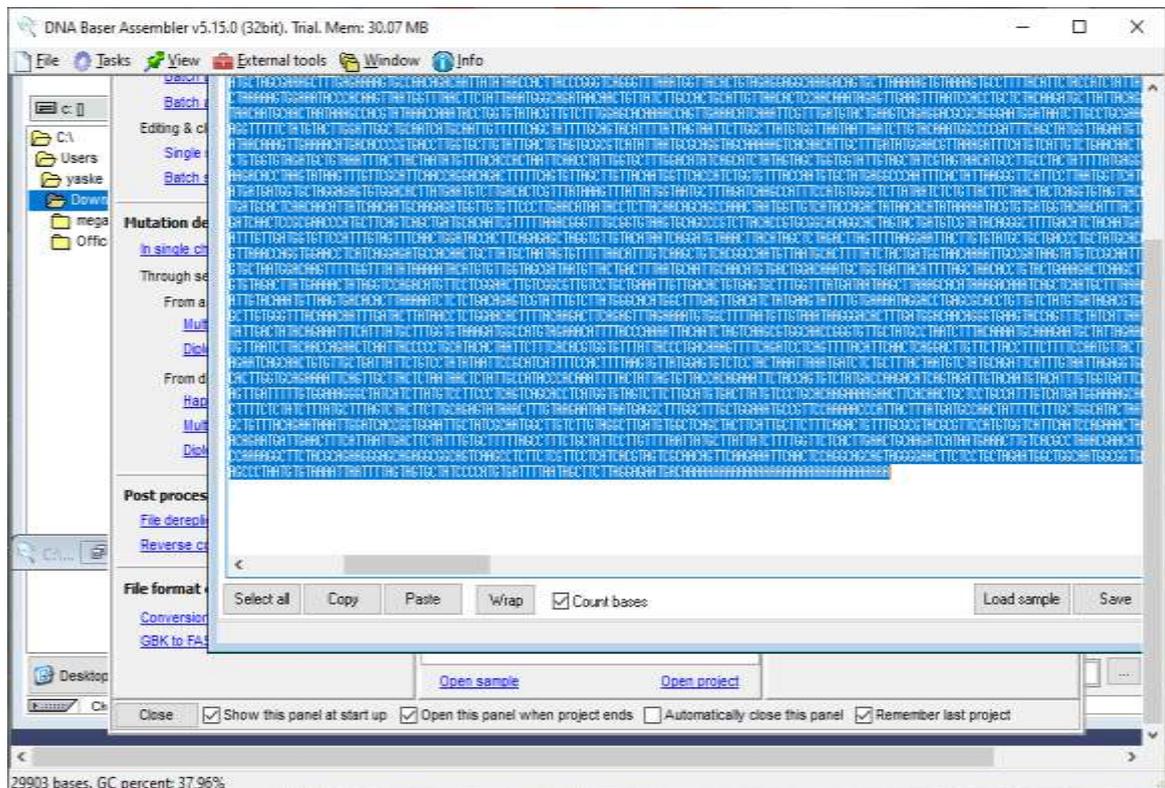
➤ La fenêtre qui apparaît ensuite permet l'introduction des séquences d'intérêt en cliquant sur Editer / Insérer une séquence vierge et en insérant le nom de pays, numéro d'accès et l'année.



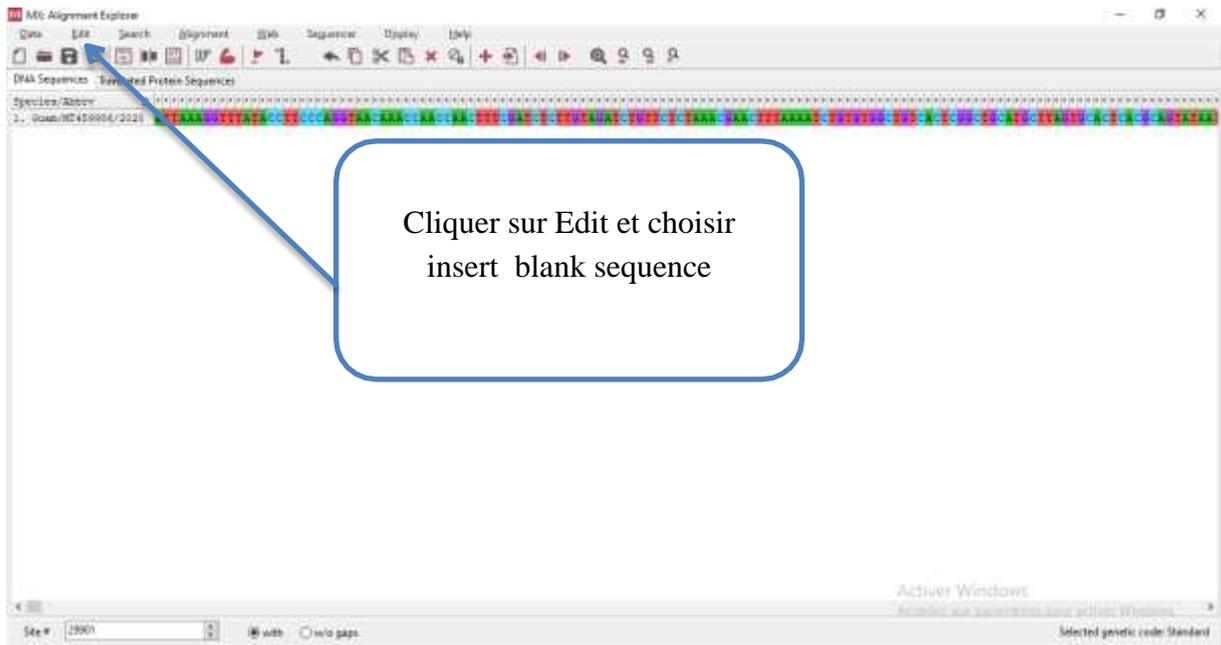
Etape 11. Cliquer sur séquence et écrire nom de pays/numéro d'accès/année.



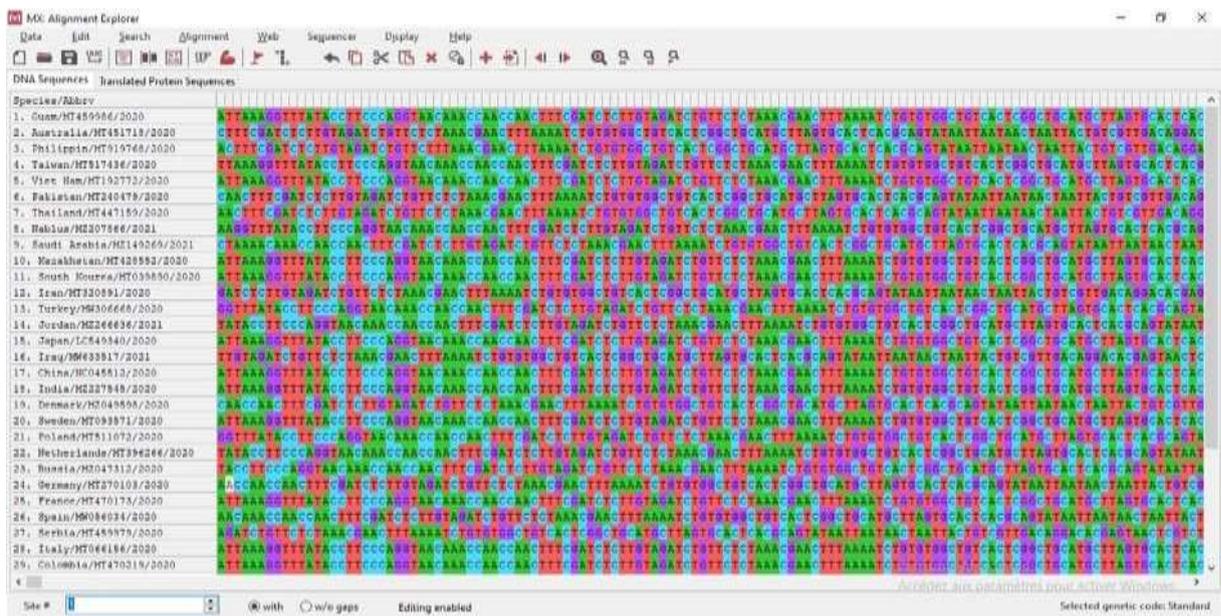
Etape 12. Une fois que le téléchargement de toutes les séquences est réalisé, les séquences sont assemblées dans DNA Baser (application installée préalablement (www.dnabaser.com)).



Etape 13. Ouvrir chaque séquence sous forme FASTA et la copier.

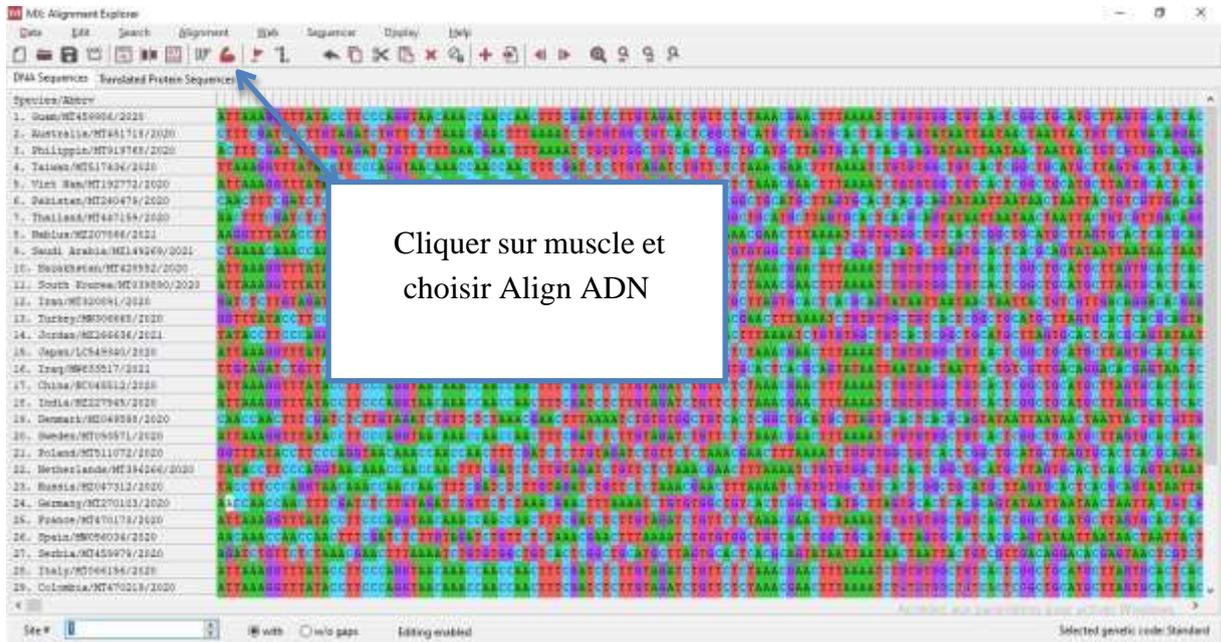


Etape 14. Insertion de 42 séquences ADN de SARS-CoV-2.

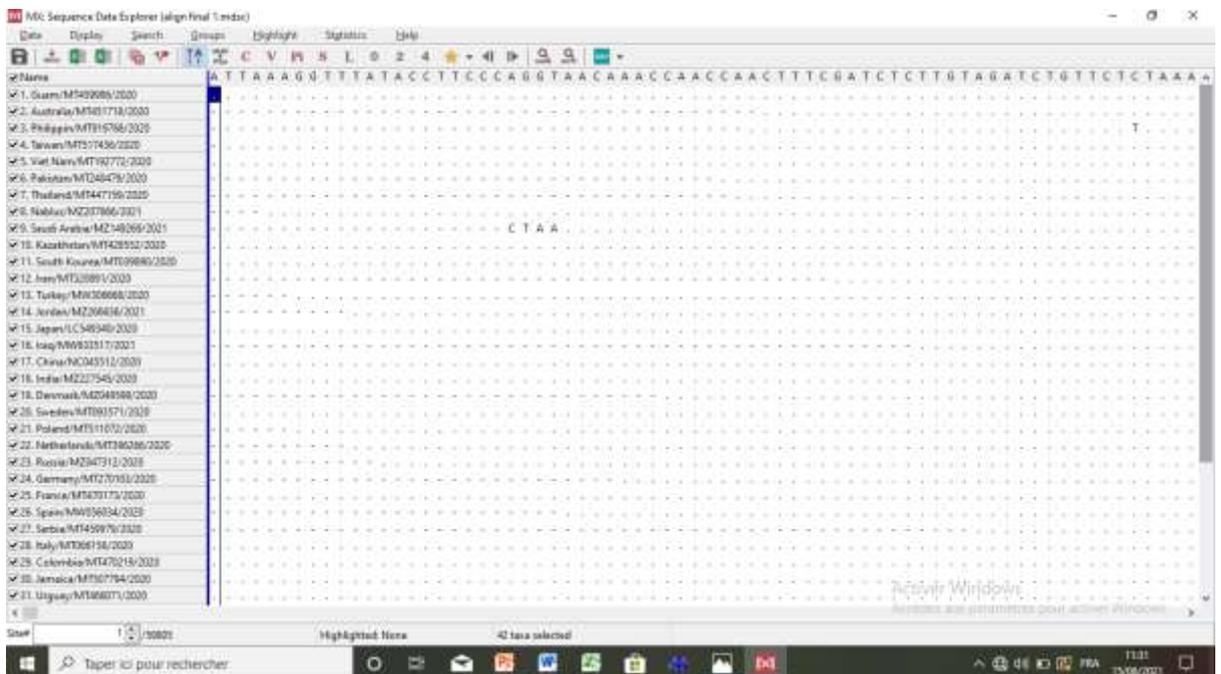


Etape 15. Résultat d'insertion des 42 échantillons d'ADN du Virus obtenus d'NCBI.

- Sélectionnez le signe muscle, cliquer « Ok » pour accepter les paramètres par défaut de align by muscle.
- L'alignement est conduit après 6h et 33 min.

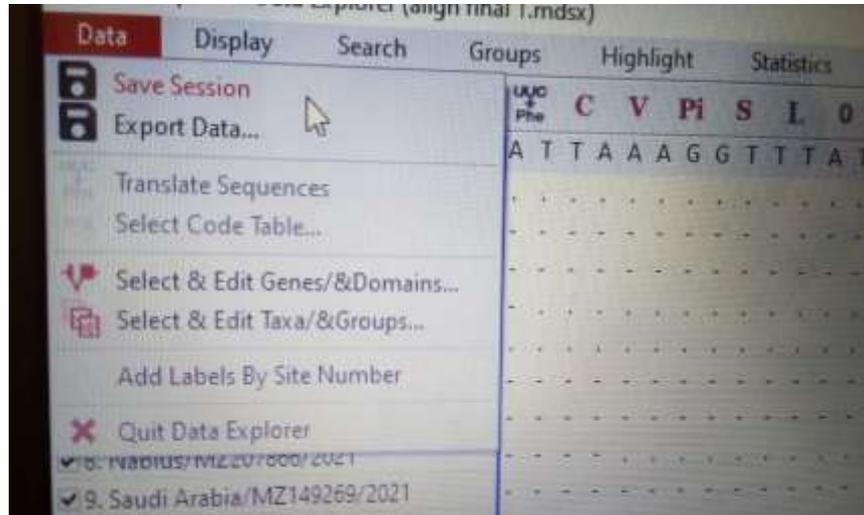


Etape 16. Sélectionnez Edit / select all pour sélectionner toutes les séquences.



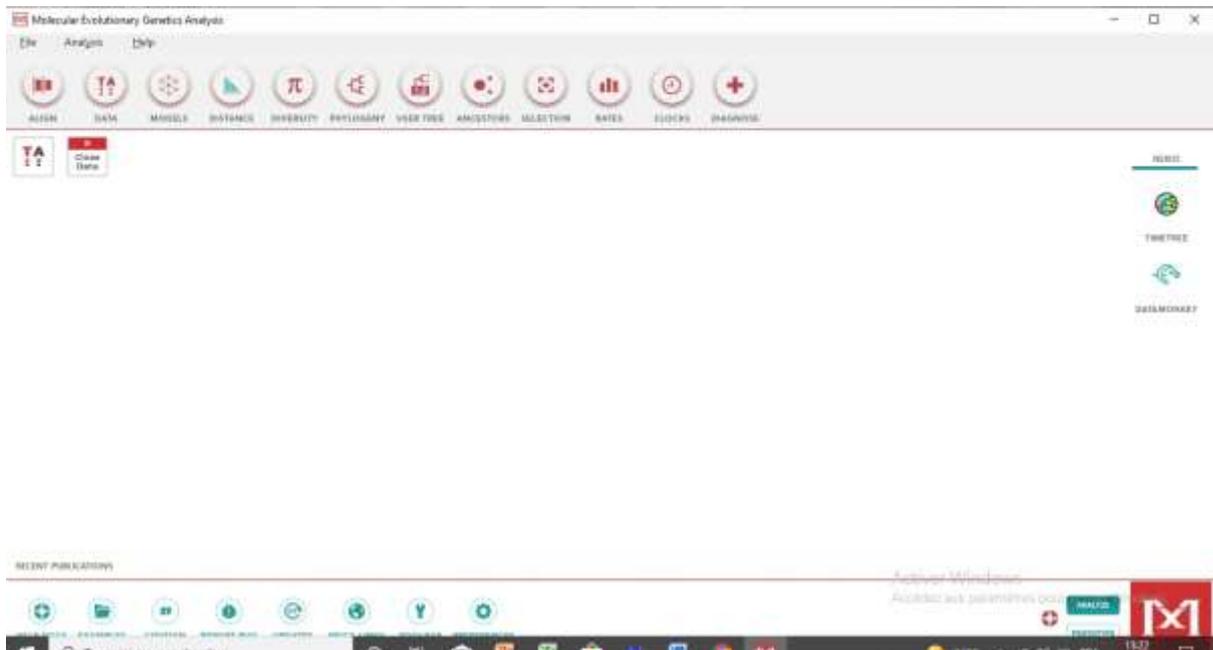
Etape 17. Résultat d'alignement

- Une fois l'alignement terminé, enregistrez la session d'alignement en cours en sélectionnant Data | Save session dans le menu principal. Donnez au fichier un nom approprié, tel que "align Final". Cela permettra à la session d'alignement en cours d'être restaurée pour une édition future.

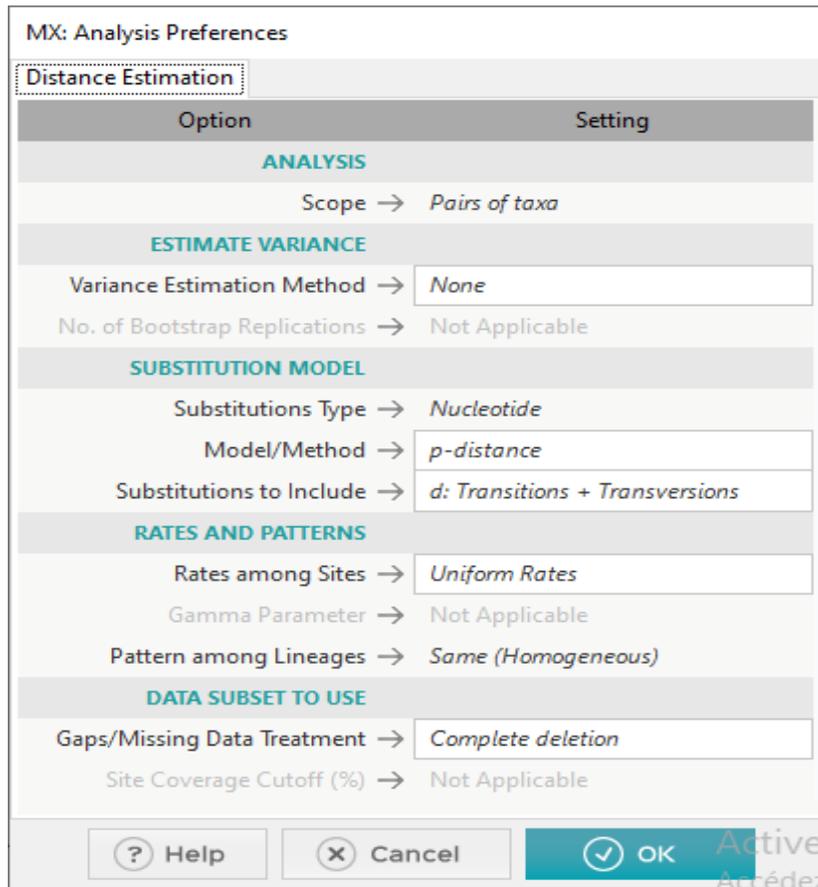


Etape 18. Enregistrement des résultats de l'alignement.

1.4.3. Estimation des distances évolutives à l'aide de la distance par paires



Etape 19. Dans la barre de lancement principale de MEGA, sélectionnez Distance |compute pairwise distance.

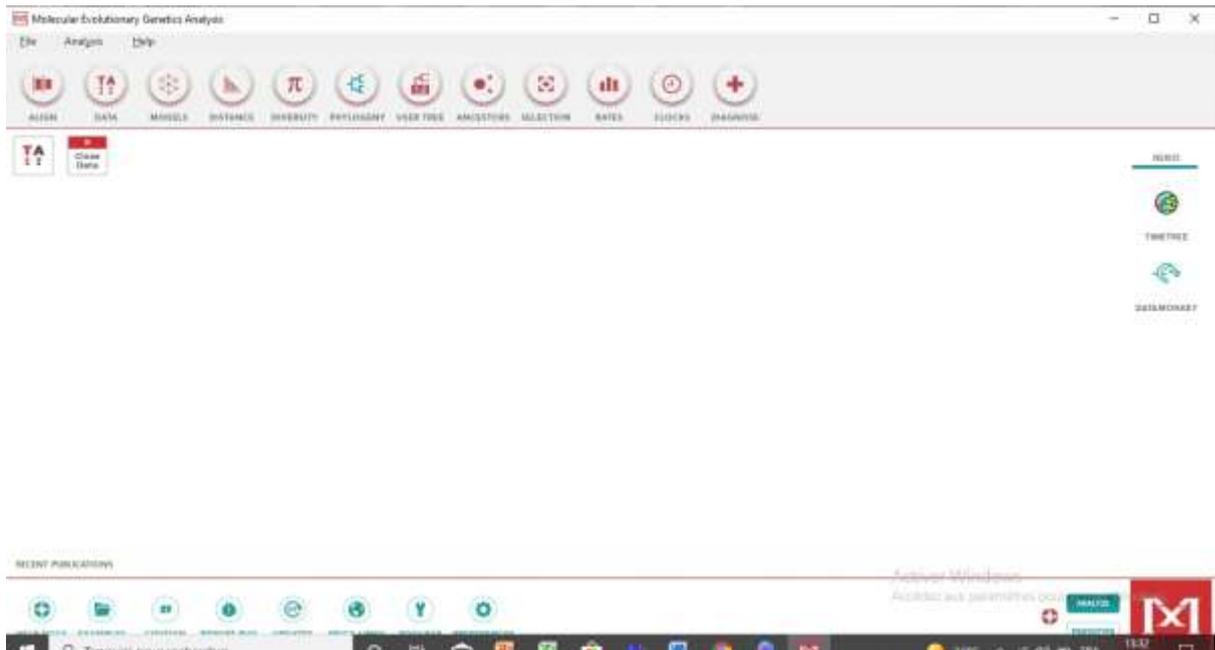


Etape 20. Cliquer sur OK pour lancer la construction de la matrice des distances.

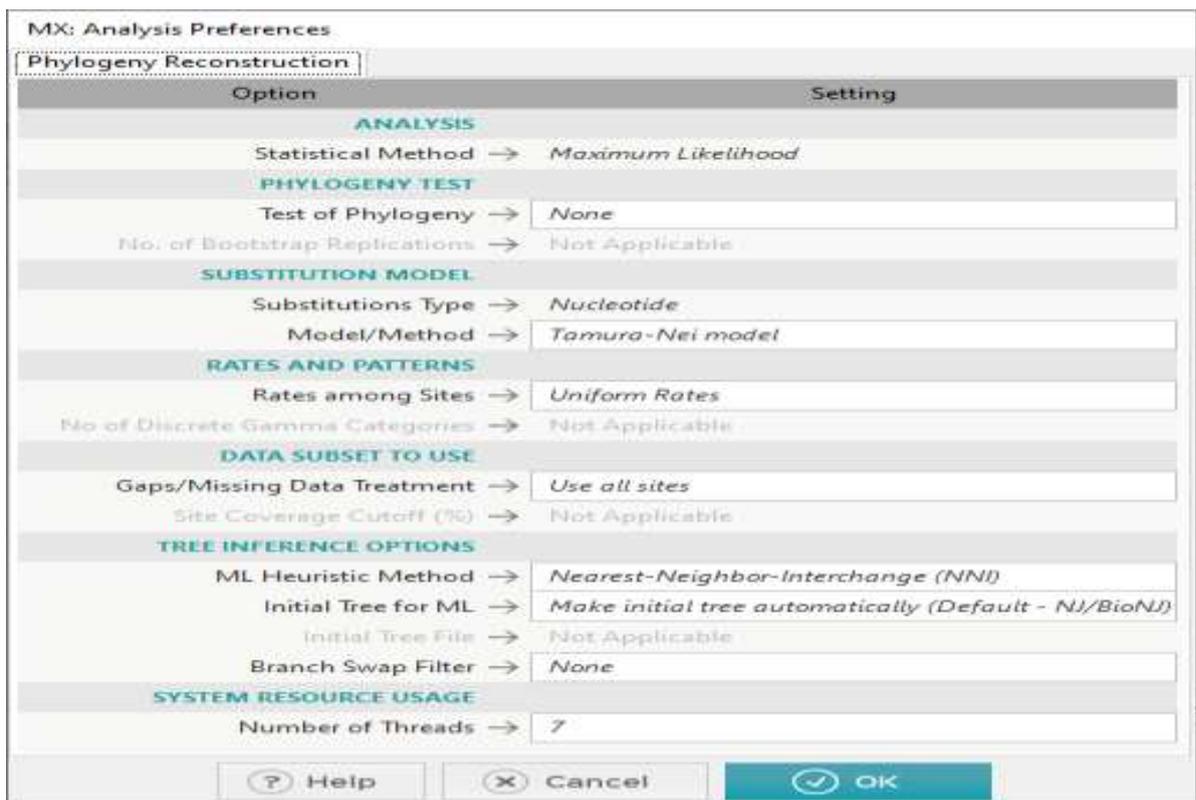
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28			
1. Guam/MT459906/2020	0.000000																														
2. Australia/MT451718/2020	0.3589901	0.000000																													
3. Philippin/MT919768/2020	0.3723807	0.3659409	0.000000																												
4. Taiwan/MT517436/2020	0.3653276	0.3726021	0.3708474	0.000000																											
5. Viet Nam/MT192772/2020	0.0001704	0.3589901	0.3724829	0.3654288	0.000000																										
6. Pakistan/MT240479/2020	0.3805915	0.3577294	0.3692630	0.3701319	0.3605574	0.000000																									
7. Thailand/MT447159/2020	0.3709496	0.3691779	0.3658046	0.3605404	0.3708815	0.3658307	0.000000																								
8. Nakhos/MZ207866/2021	0.3748509	0.3705067	0.3707622	0.3581212	0.3747998	0.3709666	0.3578827	0.000000																							
9. Saudi Arabia/MZ148266/2021	0.3601486	0.3630277	0.3716481	0.3710178	0.3601145	0.3606767	0.3754131	0.3722103	0.000000																						
10. Kazakhstan/MT428552/2020	0.0003407	0.3589730	0.3725601	0.3654809	0.0002305	0.3605745	0.3709326	0.3747658	0.3600293	0.000000																					
11. South Korea/MT038996/2020	0.0004429	0.3591434	0.3725851	0.3656002	0.0003407	0.3606426	0.3709496	0.3749020	0.3602167	0.0005111	0.000000																				
12. Iran/MT320891/2020	0.3701148	0.3745784	0.3806256	0.3709666	0.3700296	0.3721422	0.3741695	0.3704044	0.3712733	0.3700978	0.3701830	0.000000																			
13. Jordan/MZ206668/2020	0.3804722	0.3574057	0.3705748	0.3745613	0.3604041	0.3615115	0.3705407	0.3693482	0.3606256	0.3603071	0.3605745	0.3724488	0.000000																		
14. Jordan/MZ266636/2021	0.3729599	0.3733517	0.3732666	0.3579679	0.3720910	0.3715800	0.3630788	0.3624314	0.3717674	0.3728236	0.3729940	0.3696889	0.3729940	0.000000																	
15. Japan/CS49340/2020	0.0004089	0.3589730	0.3725851	0.3655320	0.0003067	0.3606085	0.3709837	0.3747998	0.3600974	0.0002044	0.0005792	0.3701659	0.3701659	0.3701659	0.000000																
16. Iraq/MW633517/2021	0.3742376	0.3705578	0.3583257	0.3707792	0.3742036	0.3715970	0.3732836	0.3722103	0.3739991	0.3741354	0.3743398	0.3606767	0.3743398	0.3743398	0.3700126	0.000000															
17. China/HK043512/2020	0.0001363	0.3589730	0.3724829	0.3654128	0.0000341	0.3605404	0.3708644	0.3747828	0.3601145	0.0002044	0.0003067	0.3700126	0.3700126	0.3700126	0.3700126	0.000000															
18. India/MZ227545/2020	0.0009370	0.3592797	0.3727895	0.3657024	0.0008348	0.3608982	0.3712392	0.3749702	0.3601997	0.0007326	0.0011074	0.3703193	0.3703193	0.3703193	0.3703193	0.000000															
19. Denmark/MZ049598/2020	0.3708304	0.3719036	0.3625848	0.3697911	0.3707452	0.3734710	0.3729258	0.3738799	0.3702341	0.3706941	0.3708985	0.3596204	0.3708985	0.3708985	0.3596204	0.3708985	0.000000														
20. Sweden/MT093571/2020	0.0003748	0.3591093	0.3725851	0.3655832	0.0002726	0.3606767	0.3709155	0.3748850	0.3602167	0.0004429	0.0004770	0.3701659	0.3701659	0.3701659	0.3701659	0.3701659	0.000000														
21. Poland/MT311072/2020	0.3605063	0.3574398	0.3706500	0.3745784	0.3604382	0.3615115	0.3705748	0.3693993	0.3606996	0.3604211	0.3606085	0.3724999	0.3724999	0.3724999	0.3701830	0.3701830	0.000000														
22. Netherlands/MT396266/2020	0.3722273	0.3736243	0.3723877	0.3561961	0.3721422	0.3724999	0.3621248	0.3610515	0.3701489	0.3720399	0.3723125	0.3701830	0.3701830	0.3701830	0.3701830	0.3701830	0.3701830	0.000000													
23. Russia/MZ047312/2020	0.3581383	0.3622270	0.3725170	0.3705237	0.3581553	0.3590923	0.3721933	0.3722103	0.3569457	0.3580701	0.3582405	0.3705918	0.3705918	0.3705918	0.3705918	0.3705918	0.3705918	0.000000													
24. Germany/MT270103/2020	0.3737095	0.3720910	0.3563154	0.3724999	0.3736754	0.3742717	0.3721592	0.3725681	0.3720059	0.3736754	0.3738458	0.3603360	0.3603360	0.3603360	0.3603360	0.3603360	0.3603360	0.3603360	0.000000												
25. France/MT470173/2020	0.0003067	0.3589560	0.3725340	0.3654639	0.0002044	0.3605574	0.3709496	0.3747828	0.3600293	0.0001022	0.0004770	0.3701148	0.3701148	0.3701148	0.3701148	0.3701148	0.3701148	0.3701148	0.000000												
26. Spain/MW036034/2020	0.3722955	0.3714437	0.3734880	0.3594501	0.3722273	0.3704726	0.3583597	0.3603189	0.3732325	0.3722273	0.3723638	0.3727895	0.3727895	0.3727895	0.3727895	0.3727895	0.3727895	0.3727895	0.3727895	0.000000											
27. Serbia/MT459979/2020	0.3536066	0.3603018	0.3751405	0.3702511	0.3536555	0.3631810	0.3714807	0.3742887	0.3583257	0.3536044	0.3536577	0.3723807	0.3723807	0.3723807	0.3723807	0.3723807	0.3723807	0.3723807	0.3723807	0.000000											
28. Italy/MT066156/2020	0.0002044	0.3589901	0.3724829	0.3654469	0.0001022	0.3605574	0.3708474	0.3748169	0.3601316	0.0002726	0.0003067	0.3700688	0.3700688	0.3700688	0.3700688	0.3700688	0.3700688	0.3700688	0.3700688	0.3700688	0.000000										

Etape 21. Résultats de l'estimation des distances évolutives (matrice des distances).

1.4.4. La construction de l'arbre phylogénétique

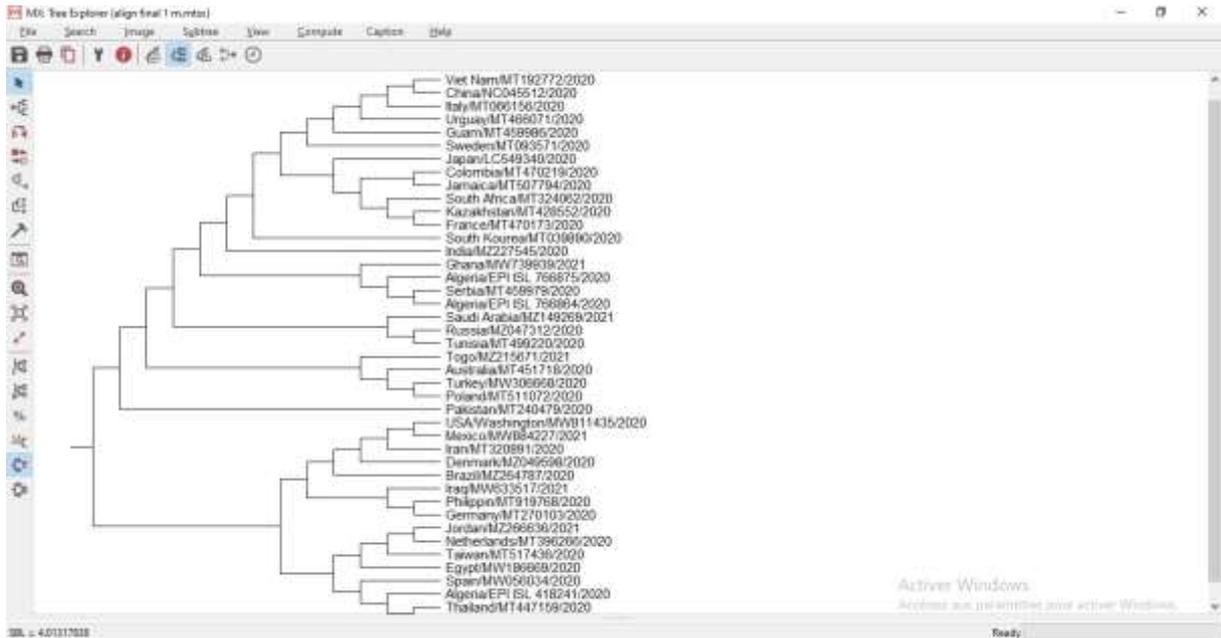


Etape 22. Sélectionner phylogeny / construct /test maximum likelihood tree



Etape 23. Cliquer OK.

➤ L'arbre phylogénétique obtenu est représenté ci-dessous.



Etape 24. L'arbre phylogénétique

*Résultats
et discussion*

2. Résultats et discussion

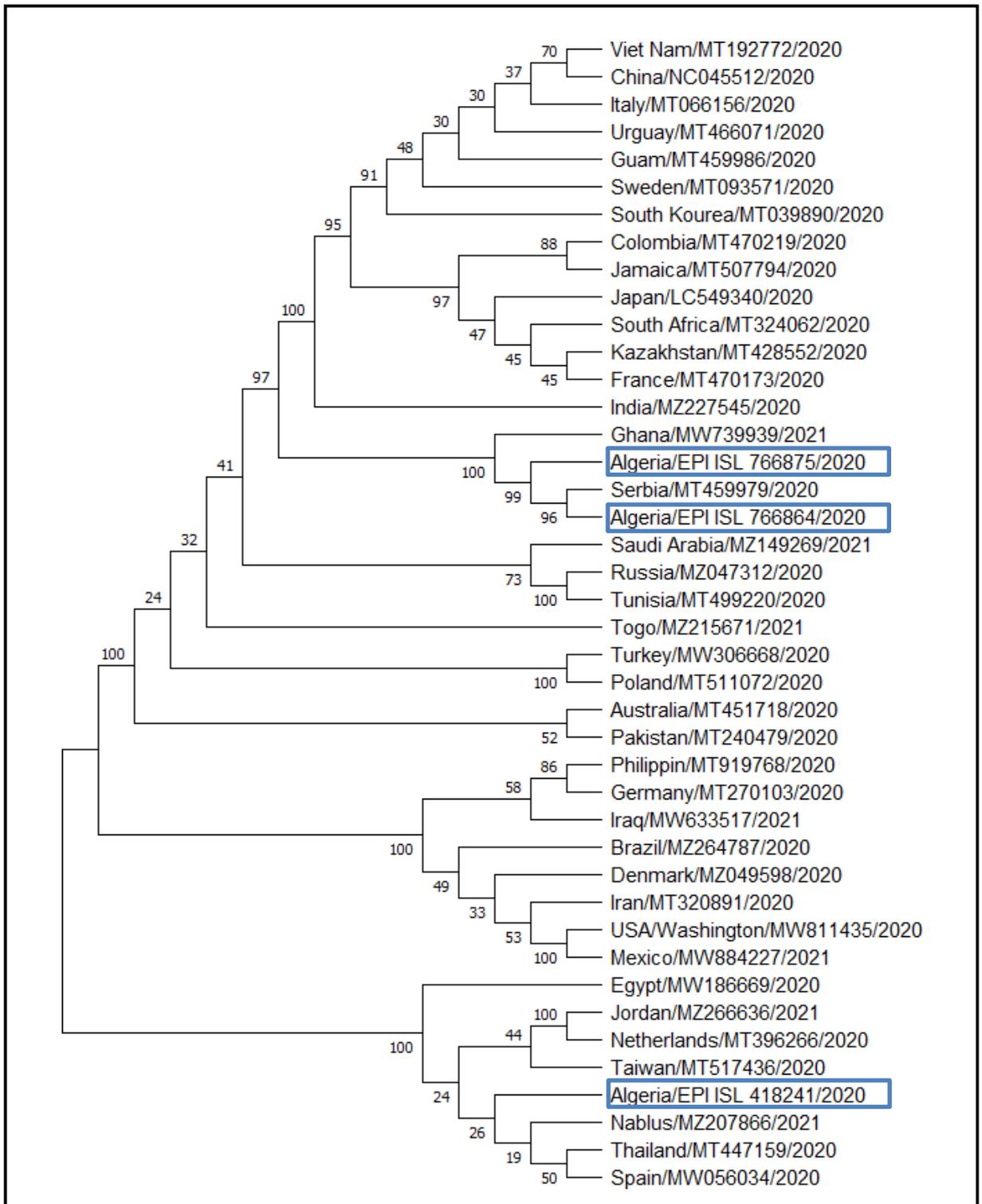


Figure 17. Arbre phylogénétique de génome complet du SARS-CoV-2 de différente région géographique. Les isolats sont représentés avec, le pays/ le numéro d'accès d'origine et l'année d'isolement.

L'histoire de l'évolution a été déduite à l'aide de la méthode NJ. Les distances évolutives ont été calculées à l'aide de la méthode de p-distance. L'analyse a porté sur 42 séquences nucléotidiques, les analyses phylogénétiques ont été menées en utilisant le logiciel MEGA X (Kumar et al., 2018).

L'arbre phylogénétique (Figure 14) a révélé la présence de 06 clades et 04 taxa isolés représentant les isolats de SARS-CoV-2 de différent pays par continent avec le numéro d'accès sur GenBank et l'année de l'identification (Tableau 5). D'abord, il est remarqué qu'il y'a une diversité génétique importante et variée par rapport à l'origine géographique. Cependant, il est constaté que les trois isolats d'Algérie sont situés dans deux clade différents et éloignées.

Tableau 5. Répartition des isolats dans les clades.

Les clades	Pays, numéro d'accès, année
Clade 1	Viet Nam/MT192772/2020
	China/NC045512/2020
	Italy/MT066156/2020
	Urguay/MT466071/2020
	Guam/MT459986/2020
	Sweedeen/MT093571/2020
	South Kourea/MT039890/2020
Clade 2	Colombia/MT470219/2020
	Jamaica/MT507794/2020
	Japan/LC549340/2020
	South Africa/MT324062/2020
	Kazakhstan/MT428552/2020
	France/MT470173/2020
Clade 3	Ghana/MW739939/2021
	Algeria/EPI ISL766875/2020
	Serbia/MT459979/2020
	Algeria/EPI ISL766864/2020
Clade 4	Saudi Arabia/MZ14926/2021
	Russia/MZ04732/2020
	Tunisia/MT499220/2020
Clade 5	Philippin/919768/2020
	Germany/MT270103/2020
	Iraq/MW633517/2021
	Brazi/MZ264787/2020
	Denmark/MZ049598/2020
	Iran/MT320891/2020

	USA/washington/MW811435/2020
	Mexico/MW884227/2021
Clade 6	Egypt/MW186669/2020
	Jordan/MZ266636/2021
	Netherlands/MT396266/2020
	Taiwan/MT517436/2020
	Algeria/EPI ISL418241/2020
	Nblus/MZ207866/2021
	Thailand/MT447159/2020
	Spain/MW056034/2020
	India/MZ227545/2020
	Togo/MZ215671/2021
Taxa 3	Turky/MW506668/2020
	Poland/MT511072/2020
Taxa 4	Australia/MT451718/2020
	Pakistan/MT240479/2020

Deux isolats Algériens (EPI ISL 766875/2020 et EPI ISL 766864/2020) situés dans un même clade (clade 3) en haut, sont regroupés avec ceux de la Serbie (MT459979) et le Ghana (MW739939) avec des distances évolutives très étroites inférieures à 0,001 (Tableau 6).

Tableau 6. Les valeurs de distance évolutive entre les pays.

	Algérie EPI ISL4181 41/02-03- 2020	Algérie ESI ISL766864/1 5-07-2020	Algérie/ EPI ISL766875 /17-06- 2020	Serbie/ MT459 979/202 0	Taiwan /MT51 7436/20 20	Ghana/ MW7399 39/2021	Arabie Saoudite/ MZ149269 /2021	Nablus/ MZ2078 66/2021	Italie/ MT0661 56/2020	France/ MT47017 3/2020
Algérie EPI ISL766864/ 15-07-2020	0.3745	-	0.00034	0.00015	0.3705	0.00009	0.3582	0.3742	0.3535	0.3534
Algérie EPI ISL766875/ 17-06-2020	0.3746	0.00034	-	0.00046	0.3703	0.0010	0.3584	0.3744	0.3535	0.3534
Algérie EPI ISL418241/ 02-03-2020	-	0.3745	0.3746	0.3745	0.3576	0.3749	0.3721	0.3580	0.370	0.3703

Le troisième isolat Algérien (EPI ISL418241) est situé dans un autre clade (clade 6) en bas regroupé avec ceux de l’Egypte (MW186669), Jordanie (MZ266636), Netherlands (MT396266), Taiwan (MT517436), Nablus (MZ207866), Thailand (MT447159) et Espagne (MW056034) avec les distances évolutives les plus faibles remarquées avec les isolats de Taiwan ($D=0,3576$) et Nablus ($D=0,3358$).

Les deux premiers isolats Algériens (EPI ISL766864, EPI ISL766875) situés dans le clade 3 sont étroitement liés entre eux avec une distance évolutive très faible de 0.00034. Alors que les troisièmes isolats (EPI ISL418241) était loin d’eux, situés dans un autre clade (clade 6) avec des distances évolutives de l’ordre de $D=0,3745$ avec le premier isolat A (EPI ISL766864) et $D=0,3746$ avec le deuxième isolat B (EPI ISL766875). Ces résultats de la phylogénie des isolats algériens montrent clairement qu’il n’y a pas une seule souche de SARS-CoV-2 en Algérie, donc différentes origines et parcours évolutifs du virus sont suggérés.

Selon Hoummadi et al (2020), en raison de la position géographique des pays d’Afrique du Nord, et en étant au plus haut niveau de contact internationale à travers le commerce, tourisme, les voyages diplomatiques, les études ou les affaires, certaines régions seraient plus touchées que d’autres par la pandémie COVID-19. Ce fait peut expliquer l’importance des cas dans ces pays. Donc, probablement la similarité génétique des isolats algériens là avec ceux de la Serbie, Taiwan, Ghana et Nablus pourrait être due à la propagation du virus par l’un de ces facteurs.

Selon Koissy (2020), Sirkeci et Mustafa, (2020), les flux migratoires de populations par le moyen des transports dans le contexte de la mondialisation, entraînent la diffusion de la maladie dans le monde entier. Une étude a été développée sur l’intensité de la migration et des voyages qui peut expliquer l’épidémie et la propagation du COVID-19 depuis son apparition fin 2019. Donc il peut être suggéré que la similarité génétique entre les isolats viraux de Ghana (MW739939) et d’Algérie (EPI ISL7668464) pourrait être due au phénomène migratoire.

Selon Jaidi (2020), Bontempi (2020), les économies africaines sont très dépendantes des échanges extérieurs, la majorité des échanges commerciaux (entre 70 et 90%) s’effectuent avec les pays du nord par le moyen des transports. Le facteur économique/commercial semble être la raison la plus plausible pouvant justifier la phase initiale de diffusion du

COVID19. Ce qui pourrait expliquer la similarité génétique entre les deux isolats de l'Algérie (EPI ISL766864 ; EPI ISL766875) avec ceux de la Serbie (MT459979) et Taiwan (MT517436).

Selon Indseth et al. (2020), COVID-19 a émergé dans un monde étroitement connecté par la population locale et internationale en mouvement, avec plus de personnes se déplaçant pour le travail, l'éducation et des raisons familiales, le tourisme et survie que jamais dans le passé. Des mouvements de population intenses, en particulier des touristes et les travailleurs d'affaires, ont été des facteurs clé de la propagation mondiale de l'épidémie.

Selon Eslami., Mahrokh (2020) et Hoummadi et al (2020), il y a des facteurs et des conditions environnementales tels que température humidité, transmission par les aliments et les emballages des aliments, l'air, La population particulièrement jeune, seraient à l'origine d'une propagation rapide de l'infection.

D'après l'analyse phylogénétique, le troisième isolat d'Algérie (EPI ISL418241) est éloigné des deux autres (EPI ISL766864, EPI ISL766875), ce qui suggère qu'il y'a deux souches différentes de SARS-CoV-2 en Algérie.

Selon Forster et al. (2020), les réseaux phylogénétiques peuvent également être utilisés avec succès pour aider à retracer les sources d'infection COVID-19, qui peuvent ensuite être mises en quarantaine pour empêcher la propagation récurrente de la maladie à l'échelle mondiale.

Conclusion

Conclusion

À notre connaissance, la présente étude l'une des premières sur la phylogénétiques comparative des souches Algériennes et avec celles d'autres pays du monde.

Nous avons rassemblé les séquences nucléotidiques du génome complet du SARS-CoV-2, isolées de différents pays du monde en utilisant les ressources bioinformatiques d'NCBI et GISAID, pour construire la matrice des distances évolutives et l'arbre phylogénétique par le logiciel MEGA X.

L'arbre phylogénétique a été réalisé à la base de 42 souches isolées de cinq continents : Asie, Afrique, Amérique, Océanie et Europe. Les résultats ont montré clairement une relation très étroite des isolats Algériens avec ceux de quatre pays géographiquement éloignés : Ghana, Taiwan, Serbie et Nablus (Palestine), d'un autre côté un rapprochement est observé entre les isolats Algériens et ceux d'Égypte, Espagne, Hollande et Jordanie.

La propagation du virus entre ces pays a comme cause plusieurs facteurs tels que l'immigration, les échanges commerciaux et le tourisme.

Par ailleurs, les isolats Algériens ont été distribués dans 2 clades distincts qui suggèrent qu'il y a des souches virales d'origine géographique et évolution différentes en Algérie.

En perspective, il faut analyser toutes les séquences génomiques Algériennes disponibles du SARS-CoV-2 pour comprendre l'origine et l'évolution des souches virales et pour détecter les mutations génétiques des isolats. Aussi, il est intéressant de faire une analyse phylogénétique à la base du gène codant pour la protéine S responsable de fixation sur les récepteurs des cellules hôte pour mieux comprendre la pathogénicité du virus. De même, il est important de faire une analyse phylogénétique à la base de la protéine S elle-même pour comprendre les interactions entre les acides aminés de son site actif et les récepteurs cellulaires.

Liste de références

Références

(A)

- Astuti, I. (2020) Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) an overview of viral structure and host response. *Diabetes Metabolic Syndrome: Clinical Research Reviews* .14(4): 407-412.

(B)

- Belouzard, S., Millet, J. K., Licitra, B. N., Whittaker, G. R. (2012) Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses*. 4(6):1011-1033.
- Bonny, V., Maillard, A., Mousseaux, C., Plaçais, L., Richier, Q. (2020) COVID-19 : physiopathologie d'une maladie à plusieurs visages. *La Revue de Médecine Interne*. 41(6) : 375-389.
- Bontempi, Elza. "Commercial exchanges instead of air pollution as possible origin of COVID-19 initial diffusion phase in Italy: more efforts are necessary to address interdisciplinary research." *Environmental research*.188 (2020).109775.
- Bontempi, E., et al. "Can commercial trade represent the main indicator of the COVID-19 diffusion due to human-to-human interactions? A comparative analysis between Italy, France, and Spain." *Environmental Research* (2021).111529.

(C)

- Corteggiani, M., Gombert, L., Pellegrini, C., Aussel, L. (2021). L'ARN polymérase COVID-19-Le talon d'Achille du SARS-CoV-2. *Médecine/Sciences*.37(3):288-292.
- Couvreur, P., Louvard D. (2021) "COVID-19 and drugs: Pathophysiology and Therapeutic Approaches. *Comptes Rendus. Biologies*.1-16.

(D)

- Demongeot, J., Flet-Berliac, Y., Seligmann, H. (2020). La température diminue les paramètres de la dynamique de propagation des nouveaux cas de Covid. 9(5).94 :1-9.
- Didier R. (2020), *Épidémies vrais dangers et fausses alerts de la grippe aviaire*; Editions Michel Lafon.1-96.

(E)

- Eslami, Hadi, and Mahrokh Jalili. "The role of environmental factors to transmission of SARS-CoV-2 (COVID-19)." *Amb Express*.10 (2020):1-8.

(F)

- Forster, Peter, et al. "Phylogenetic network analysis of SARS-CoV-2 genomes." *Proceedings of the National Academy of Sciences*.117.17 (2020) : 9241-9243).

(G)

- Gadhave, K., Kumar, P., Kumar, A., Bhardwaj, T., Garg, N., Giri, R. (2020). NSP 11 of SARS-CoV-2 is an Intrinsically Disordered Protein. *bioRxiv*.1-34.
- Ge, H., Wang, X., Yuan, X., Xiao, G., Wang, C., Deng, T., Yuan, Q., Xiao, X., 2020. The epidemiology and clinical information about COVID-19. *Eur. J. Clin. Microbiol. Infect. Dis.* 39 : 1011–1019. <https://doi.org/10.1007/s10096-020-03874-z>.
- Gligorov, J., Bachelot, T., Pierga, J. Y., Antoine, E. C., Balleyguier, C., Barranger, E., Delaloge, S. (2020). COVID-19 et personnes suivies pour un cancer du sein: recommandations françaises pour la pratique clinique de Nice-St Paul de Vence, en collaboration avec le Collège Nationale des Gynécologues et Obstétriciens Français (CNGOF), la Société d’Imagerie de la Femme (SIFEM), la Société Française de Chirurgie Oncologique (SFCO), la Société Française de Sénologie et Pathologie Mammaire (SFSPM) et le French Breast Cancer Intergroup-UNICANCER (UCBG). *Bulletin du Cancer*.107(5):528.
- Gorbalenya, A. E., Baker, S. C., Baric, R., Groot, R. J. D., Drosten, C., Gulyaeva, A. A., Ziebuhr, J. (2020). Severe acute respiratory syndrome-related coronavirus: The species and its viruses—a statement of the Coronavirus Study Group.1-15.
- Gorla, U. S., Rao, G. K. (2020). SARS-CoV-2: The Prominent Role of Non-structural Proteins (NSPS) in COVID-19. *Indian Journal of Pharmaceutical Education and Research*. S381-S389.

- Graham, R. L., Sparks, J. S., Eckerle, L. D., Sims, A. C., Denison, M. R. (2008). SARS coronavirus replicase proteins in pathogenesis. *Virus research*. 133(1) : 88-100.

(H)

- Harrison, A. G., Lin, T., Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in immunology*.41(12) :1100-1115.
- He, F., Deng, Y., Li, W. (2020). Coronavirus disease 2019 : What we know ? *Journal of Medical Virology*.92(7) :719-725.
- Hoummadi, L., Hafid, J., Machraoui, S., Admou, B. (2020). Jusqu' où l' Afrique peut- elle limiter l' impact de la pandémie de COVID-19? *Revue d'Épidémiologie et de Santé Publique*.68(5):302-305.
- Huang, Yuan, et al. "Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19." *Acta Pharmacologica Sinica* 41.9 (2020): 1141-1149.

(I)

- Indseth, Thor, et al. "COVID-19 among immigrants in Norway, notified infections, related hospitalizations and associated mortality: A register-based study." *Scandinavian Journal of Public Health* (2020): 1403494820984026.

(J)

- Jaïdi, L. (2020), La Covid-19 en Afrique: Etat des lieux et perspectives de sortie de crise. *Rapport annuel sur l'économie de l'Afrique*, Ed. Policy Center for the New South.19-337.
- Jiang, S., Hillyer, C., Du, L. (2020). Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends in immunology*. 41(5) :355-359.
- Juckel, Dylan, Jean Dubuisson, and Sandrine Belouzard. "Les coronavirus, ennemis incertains." *Médecine/Sciences*.36.6-7 (2020) : 633-641.
- Jungreis, I., Sealfon, R., Kellis, M. (2021). SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. *Nature communications*. 12(1) : 1-20.

(K)

- Kandeel, M., Ibrahim, A., Fayez, M., Al-Nazawi, M . (2020). From SARS and MERS CoVs to SARS-CoV- 2: Moving toward more biased codon usage in viral structural and nonstructural genes. *Journal of medical virology*.92(6):660-666.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*.35(6): 1547.
- Kumar, S., Nyodu, R., Maurya, V. K., Saxena, S. K. (2020). Morphology, genome organization, replication, and pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In *Coronavirus Disease 2019 (COVID-19)* (pp. 23-31). Springer, Singapore.

(L)

- Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B. T., Chaillon, A. (2020). Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV- 2. *Journal of medical virology*.92(6):602-611.
- Liu, Ding Xiang, et al. "Accessory proteins of SARS-CoV and other coronaviruses." *Antiviral research*. 109 (2014): 97-109.
- Llanes, A., Restrepo, C. M., Caballero, Z., Rajeev, S., Kennedy, M. A., Lleonart, R. (2020). Betacoronavirus genomes: how genomic information has been used to Deal with past outbreaks and the COVID-19 pandemic. *International journal of molecular sciences*. 21(12) :4546.
- Ludwig, S., Zarbock, A. (2020). Coronaviruses and SARS-CoV-2: a brief overview. *Anesthesia and analgesia*.131 (1):93-96.

(M)

- Machhi, J., Herskovitz, J., Senan, A. M., Dutta, D., Nath, B., Oleynikov, M. D., Kevadiya, B. D. (2020). The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections. *Journal of Neuroimmune Pharmacology*.1-28.

- Malik, Yashpal Singh, et al. "Emerging novel coronavirus (2019-nCoV) current scenario, evolutionary perspective based on genome analysis and recent developments." *Veterinary quarterly*. 40.1 (2020) : 68-76.
- McBride, Ruth, and Burtram C. Fielding (2012). "The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis." *Viruses* .4(11) : 2902-2923.
- Miao, Zhichao, et al. "Secondary structure of the SARS-CoV-2 5'-UTR." *RNA biology* (2020) : 1-10.
- Michel, C. J., Mayer, C., Poch, O., Thompson, J. D. (2020). Characterization of accessory genes in coronavirus genomes. *Virology journal*.17(1):1-13.
- Mondiale de la Santé, O. (2021). Séquençage génomique du SARS-CoV-2 à des fins de santé publique : orientations provisoires, 8 janvier 2021 (No. WHO/2019-nCoV/genomic_sequencing/2021.1). Organisation mondiale de la Santé.

(N)

- Naqvi, Ahmad Abu Turab, et al. "Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach." *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* (2020) : 165878.

(O)

- Olotu, F. A., Omolabi, K. F., Soliman, M. E. (2020). Piece of the Puzzle: remdesivir disassemble the multimeric SARS-CoV-2 RNA-dependent RNA polymerase non-structural proteins (RdRp-NSPs) complex. 1-24.

(P)

- Pathak, S. K. (2020). General Details of Structural Proteins of Coronaviruses with Special Reference of SARS-COV-2 or COVID-19. *Bull. Env. Pharmacol. Life Sci.* 9 :34-38.
- Perdikari, Theodora Myrto, et al. "SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins." *BioRxiv* (2020). [10.1101/2020.06.09.141101](https://doi.org/10.1101/2020.06.09.141101) .
- Pereira, Filipe. "Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene." *Infection, Genetics and Evolution*. 85 (2020) : 104525.

- Pruitt, K. D., Tatusova, T., Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): à curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*.33(suppl_1) : D501-D504.

(R)

- Raj, R. (2021). Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochemistry and biophysics reports*. 25 :100847.
- Rajnik, M., Cascella, M., Cuomo, A., Dulebohn, S. C., Di Napoli, R. (2021). Features, Evaluation, and Treatment of Coronavirus (COVID-19). *Uniformed Services University of the Health Sciences*. 1-21
- Rousseau, A., Fenolland, J. R., Labetoulle, M. (2020). SARS-CoV-2, COVID-19 et œil : le point sur les données publiées. *Journal Français d'Ophtalmologie*.43 :642-652

(S)

- Sallard, E., Halloy, J., Casane, D., van Helden, J., Decroly, É. (2020). Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. *médecine/sciences*.36(8-9):783-796.
- Santerre, M., Arjona, S. P., Allen, C. N., Shcherbik, N., Sawaya, B. E. (2020). Why do SARS-CoV-2 NSPs rush to the ER. *Journal of Neurology*.1-10.
- Segondy, M. (2020). Les Coronavirus humains. *Revue Francophone des Laboratoires*.2020(526) :32-39.
- Sharif, N., Sarkar, M. K., Ahmed, S. N., Ferdous, R. N., Nobel, N. U., Parvez, A. K., Dey, S. K. (2021). Environmental correlation and epidemiologic analysis of COVID-19 pandemic in ten regions in five continents. *Heliyon*.7(3): e06576.
- Shereen, Muhammad Adnan, et al. "COVID-19 infection: Origin, transmission, and characteristics of humancoronaviruses." *Journal of advancedresearch*. 24 (2020): 91-98.

- Sirkeci, Ibrahim, and Mustafa Murat Yucesahin. "Coronavirus and migration: Analysis of human mobility and the spread of COVID-19." *Migration Letters*. 17.2 (2020): 379-398.

(T)

- Tabibzadeh, Alireza, et al. "SARS-CoV-2 molecular and phylogenetic analysis in COVID-19 patients : à preliminary report from Iran." *Infection, Genetics and Evolution*. 84 (2020) : 104387.
- Thomas, Sunil. "The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter semiSWEET." *Pathogens and Immunity*.5.1 (2020): 342.
- Triggler, Chris R., et al. "A Comprehensive Review of Viral Characteristics, Transmission, Pathophysiology, Immune Response, and Management of SARS-CoV-2 and COVID-19 as a Basis for Controlling the Pandemic." *Frontiers in Immunology*.12 (2021) : 338.

(V)

- Van Nho, J. T., Pardo, E. (2020). Complications cardiaques de la COVID-19 en réanimation. *Le Praticien en anesthésie reanimation*.24(4):212-217.

(W)

- Wabalo, E. K., Dubiwak, A. D., Gizaw, T. S., Kotu, U. G. (2020). Role of structural and functional proteins of SARS-COV-2. *GSC Biological and Pharmaceutical Sciences*.12(3):117-129.
- Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., Jiang, T., 2020. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe*. 27:325–328.<https://doi.org/10.1016/j.chom.2020.02.001>

(X)

- Xu, J., Zhao, S., Teng, T., Abdalla, A. E., Zhu, W., Xie, L., Guo, X. (2020). Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses*. 12(2):244.

(Y)

- Yahia, F., Zakhama, L., Abdelaziz, A. B., PRP2S4, R. M. (2020). COVID-19 et Maladies cardiovasculaires. Etude scoping review COVID-19 and Cardiovascular diseases. Scoping review study. *La Tunisie medicale*.98(04):283-294.
- Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Lou, Z. (2020). Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nature communications*.11(1) :1-6.
- Ye, Qiaozhen, et al. "Architecture and self-assembly of the SARS- CoV- 2 nucleocapsid protein." *Protein Science*. 29.9 (2020): 1890-1901.
- Yoshimoto, F. K. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *The protein journal*. 39:198-216.

(Z)

- Zinzula, Luca. "Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2." *Biochemical and Biophysical Research Communications*. 538 (2021) : 116-124.
- <https://www.rcsb.org/>

Résumé

La pandémie COVID-19 devient un problème de santé publique en Algérie et dans le monde. Dans la présente étude 42 séquences du génome complet de SARS-CoV-2 de différentes géographies dans monde ont été recueillies sur les bases de données NCBI et GISAID. Les séquences nucléotidiques ont fait l'objet d'un alignement, une estimation des distances évolutives et la construction d'un dendrogramme par l'outil bioinformatique MEGA X. Les résultats ont montré une variabilité génétique importante du SARS-CoV-2 dans le monde.

Mots clés : Phylogénie, SARS-CoV-2, variabilité génétique, distances évolutives

Summary

The COVID-19 pandemic is becoming a public health problem in Algeria and around the world. In the present study 42 complete genome sequences of SARS-CoV-2 from different geographies around the world were collected on the NCBI and GISAID databases. The nucleotide sequences were the subject of an alignment, an estimation of the evolutionary distances and the construction of a dendrogram by the bioinformatics tool MEGA X. The results showed a significant genetic variability of SARS-CoV-2 in the world.

Key words: Phylogeny, SARS-CoV-2, genetic variability, evolutionary distances

ملخص

أصبحت جائحة COVID-19 مشكلة صحية عامة في الجزائر وحول العالم. في هذه الدراسة تم جمع 42 تسلسل جينومي كامل لـ SARS-CoV-2 من مناطق جغرافية مختلفة في العالم في قاعدتي بيانات NCBI وGISAID. كانت متواليات النيكلويدات موضوع محاذاة وتقدير للمسافات التطورية وبناء مخطط شجرة بواسطة أداة المعلوماتية الحيوية. أظهرت النتائج تبايناً جينياً كبيراً لـ SARS-CoV-2 حول العالم.

الكلمات المفتاحية: علم الوراثة، SARS-CoV-2، التباين الجيني، المسافات التطورية