

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique
Université de Blida 1



Faculté des sciences de la nature et de la vie
Département biologie

Mémoire de fin d'étude

En vue de l'obtention du diplôme de
Master en Sciences de la nature et de la vie

Option : Génétique

Thème

**Etude des différentes mutations détectées chez les gènes de
coronavirus en utilisant des outils de bio-informatique.**

Présenté par :

SEDDAOUI Rania

LATRACH Rabia

Devant le jury :

Mme AMOKRANE.

M.A.A USDB 1

Présidente

Mme ABDULHUSSEIN A.

M.C.A USDB 1

Examinatrice

Mme DEROUICHE L.

M.C.A ESSAIA

Promotrice

Mme ZATRA Y.

M.C.A USDB 1

Co-promoteur

Promotion: 2021/2022



Remerciements

*D'abord nous remercions Dieu le tout puissant de
Nous avoir donné courage, santé, souffle et patience pour
Accomplir ce travail.*

*Nous remercions vivement **Mme DEROUICHEL.***

*Pour sa disponibilité et toutes les connaissances
Qu'elle nous a transmises, pour la qualité de son encadrement
Exceptionnel et les*

*Interventions enrichissantes et encourageantes qu'elle
Nous a accordées au cours de ce travail.*

Nos vifs remerciements vont également aux membres du jury :

***Madame ABDULHUSSEIN A.** pour l'intérêt qu'elle a porté à notre
recherche en acceptant*


D'examiner notre travail et de l'enrichir par ses propositions.

***Madame AMOKRANE A.** de nous avoir fait le plaisir de présider ce
jury.*

Nous sommes très honorés de leur présence dans ce jury.

*Nos profonds remerciements vont également à la Maitre assistante
au département de biologie **Madame ZATRA Y.** qui nous a permis de
réaliser*

Ce travail et de nous avoir porté aide.



Dédicaces

Du profond de mon cœur, je dédie ce travail à tous ceux qui me sont chers,

Je remercie ALLAH

*Tout puissant qui nous guide dans notre vie, qui nous aide dans notre chemin, je Lui dois ce que je suis devenue ce travail est de Lui, grâce à Lui et vers Lui.
Louanges et remerciements pour Votre clémence et miséricorde.*

A mon cher père Abderrahmane

*Les mots me manquent pour décrire le formidable Homme et père que tu es, tu es toujours présent et bienveillant, tu es mon pilier.
Merci beaucoup pour les ailes que tu m'as données, pour les sacrifices que tu as faits, pour tout ton amour et ta tendresse.
Merci pour tout, j'espère pouvoir t'apporter autant dans les prochaines années et te combler de bonheur et que tu sois fier de moi. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur. Je t'aime très fort mon roi.*

A ma très chère maman Fethia

*La prunelle de mes yeux, la meilleure maman du monde, tu avais des rêves mais tu les as souvent abandonnés pour me permettre de vivre les miens. Merci d'être présente dans ma vie parce que sans toi je ne serais rien.
Maman chérie je ne pourrais jamais décrire l'amour que je ressens envers toi et mon père.
Ce travail n'aurait jamais vu le jour sans ton soutien, ta patience, tes sacrifices et tes encouragements continus, je te dédie ce travail en témoignage de mon amour et ma profonde reconnaissance. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur. Je t'aime très fort Maman chérie.*

A mon très cher frère Kamel et très chère sœur Nihad

J'ai de la chance de vous avoir comme grand frère et petite sœur je sais que vous serez toujours là pour moi et me rendre heureuse. Je ne remercierai jamais Dieu assez pour votre présence dans ma vie, tu es mon frère protecteur et mon ange gardien, tu es ma sœur ma source de bonheur et de motivation. Que nos liens restent toujours solides et que Dieu nous apporte bonheur et nous aide à réaliser tous nos rêves ensemble. Que Dieu vous procure santé, réussite et vous préserve du mal. Je vous aime.

***A mon fiancé Lyes**, je te remercie pour ton soutien moral, pour tes encouragements et l'aide que tu m'as toujours accordé.*

*Une spéciale dédicace pour mon binôme **Rabia** et sa famille, je te souhaite beaucoup de succès dans ta vie personnelle et professionnelle.*

A mes chères amies Meroua et profty Nassira

Je ne peux pas trouver les mots exacts et sincères pour vous exprimer mon affection et mes pensées, vous êtes pour moi des sœurs et des amies sur qui je peux compter.

A ma chère famille Seddaoui et Amirouchi

En fin, je le dédie chaleureusement à tous ceux qui me connaissent et m'aiment.

RANIA

Dédicaces

Je dédie ce travail

*Aux êtres les plus chers de ma vie **Maman***

*Et mon cher **Papa***

Qui m'ont soutenu et encouragé durant toutes ses années

Qu'ils trouvent ici le témoignage de ma profonde reconnaissance

A mes frères qui ont partagé avec moi tous les moments d'émotion lors de la

Réalisation de ce travail

*A toutes mes **amies** que je considère comme des sœurs et à qui je souhaite plus*

de succès

*Sans oublier mon binôme **Rania** pour son soutien moral, sa patience son bon*

humour et sa compréhension

Tout au long de ce projet.

RABIA

Listes des tableaux et figures

Liste des tableaux

N° de tableau	Titre	Pages
01	Caractéristiques des gènes exprimés par le SARS-CoV-2 (Astrid et al., 2020)	8
02	Caractéristiques cliniques et virologiques des principaux variants du SARS-COV (Deluzarche et al., 2020).	15
03	Séquences nucléotidique de différentes variants de Coronavirus extraites de la GenBank.	42
04	Résultats des arbres phylogénétiques obtenus par MEGA des gènes de SARS Cov-2.	45
05	Caractéristiques des différentes mutations que nous avons présentées dans l'arbre s	50
06	Caractéristiques des différentes mutations que nous avons présenté dans l'arbre du gène ORF1ab.	55
07	Caractéristiques des différentes mutations que nous avons obtenu dans l'arbre de Nucléocapside N.	60
08	Caractéristiques des différentes mutations que nous avons présentées dans l'arbre de gène ORF3a.	63

Liste des figures

N° de Figure	Titre	Page
1	Schéma représentatif de la classification de Coronaviridae (Tang et al., 2020).	3
2	Structure schématisée du SARS-CoV-2 (Gautheret 2020).	5
3	Le domaine RBD (receptor Binding domain) est la région de la protéine spike entrant en contact avec le récepteur cellulaire ACE2 qui sert de porte d'entrée au SARS-CoV-2 dans les cellules qu'il infecte (Gazzali et al., 2020).	5
4	Présentation schématique de l'organisation du génome du SRAS-CoV-2, des ARNm sous-génomiques canoniques et de la structure du virion (Meriadeg et al., 2020).	6
5	Risque de transmission du SARS-CoV-2 dans différentes situations. En vert : risque faible. En orange : risque modéré. En rouge : risque fort (Jones et al., 2020).	8
6	La mutation D614G au niveau de la protéine S du SARS-COV-2 (Zhang Y et al., 2021).	9
7	Les mutations clés présentes dans le génome du variant Anglais (Segondy et al., 2020).	11
8	Les mutations observées dans les variants « Alpha, Béta, Gamma (Zhou W, al., 2020).	12
9	Exemple d'un arbre phylogénétique (Schmidt, 2003).	21
10	Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné (Schmidt, 2003).	21
11	Diverses catégories de groupes taxonomiques. Les taxa X, Y, Z forment un groupe monophylétique. Les taxa 1, 2 et 3 forment un groupe paraphylétique. Les taxa A, B et C forment un groupe polyphylétique (Tourasse, 1992).	22
12	Page d'accueil d'NCBI (Personnelle 2022).	25
13	Saisie des mots-clés dans la GenBank (Personnelle 2022).	25
14	Filtrage des résultats de recherche (Personnelle 2022).	26

15	Sélection du résultat de recherche voulu (Personnelle 2022) .	26
16	Page du génome complet du SARS-COV-2 sur le site NCBI (Personnelle 2022) .	27
17	Identification NCBI de la séquence du gène (Personnelle 2022) .	27
18	Le choix de séquences dans la GenBank (Personnelle 2022) .	28
19	Apparition de la séquence au format FASTA (Personnelle 2022) .	28
20	Téléchargement de la séquence du gène SARS-CoV 2 sous format FASTA (Personnelle 2022) .	29
21	La page principale du programme Fabox (Personnelle 2022) .	30
22	La conversation des séquences sous forme FASTA par l'utilisation de Fabox (Personnelle 2022) .	31
23	Alignement trimmer (la forme FASTA) (Personnelle 2022) .	31
24	Page principale de web site BLAST (Personnelle 2022) .	32
25	Les paramètres de BLAST (Personnelle 2022) .	33
26	Les paramètres d'algorithme de BLAST (Personnelle 2022) .	33
27	Lecteur du détail des résultats d'alignement (Personnelle 2022) .	34
28	Résumé graphique (Personnelle 2022) .	34
29	Etape de téléchargement des similarités forme FASTA (Personnelle 2022) .	35
30	La fenêtre principale de MEGA11 (Personnelle 2022) .	36
31	Ouvrir le programme de MEGA11. Cliquer sur Align et sélection de Edit /Build Alignment (Personnelle 2022) .	36
32	Choisir create a new alignment et cliquer sur OK (Personnelle 2022) .	37
33	Cliquer sur DNA dans MEGA11 (Personnelle 2022) .	37
34	Cliquer sur séquence et écrire Cov, nom de pays, le variant dans MEGA11 (Personnelle 2022) .	37
35	Insertion de séquences ADN de SARS-CoV-2 dans MEGA (Personnelle 2022) .	38
36	Résultat d'insertion des échantillons d'ADN du Virus SARS-CoV-2 obtenus d'NCBI dans MEGA11 (Personnelle 2022) .	38
37	Sélectionner Edit / select all pour sélectionner toutes les séquences sur MEGA11 (Personnelle 2022) .	39
38	Enregistrement des résultats de l'alignement (Personnelle 2022) .	39
39	Barre de lancement principale de MEGA11 (Personnelle 2022) .	40
40	Les paramètres utilisés pour la reconstruction phylogénétique (Personnelle 2022) .	41

41	Organisation génomique du SARS-CoV-2 (Segondy 2020).	47
42	Arbre phylogénétique de gène S représentant différentes variantes de Coronavirus.	48
43	Arbre phylogénétique de gène ORF1ab représentant différentes variantes de Coronavirus.	53
44	Arbre phylogénétique de nucléocapside N représentant différentes variantes de Coronavirus.	58
45	Arbre phylogénétique de gène ORF3a représentant différentes variantes de Coronavirus.	62
46	Une courbe montrant l'évolution du Corona virus en Algérie de 2020 à 2022 (OMS 2022).	78

Liste d'abréviations

- **ACE2** : Enzyme de conversion de l'angiotensine 2.
- **ADN**: Acide Désoxyribonucléique.
- **ARN**: Acide Ribo Nucléique.
- **Blast**: Basic Local Alignment Search Tool.
- **Bootstrap** : Test de robustesse des phylogrammes.
- **COVAX**: Facility – COVID-19 Vaccine Global Access Facility.
- **DNA** : Deoxyribonucleic Acid.
- **DnaSP**: DNA Séquence Polymorphisme.
- **MEGA**: Molecular Evolutionary Genetics Analysis.
- **NCBI** : Centre for Biotechnology Information.
- **PHYLIP**: PHYLogeny Inference Package.
- **RBD** : Receptor Binding Domaine.
- **SARS-CoV-2** : Coronavirus du Syndrome Respiratoire Aigu Sévère 2
- **OMS** : Organisation Mondial de la Santé.
- **NSP**: Non Structurel Protéine.
- **RT-PCR**: Reverse Transcriptase PCR- Polymerase Chain Reaction.
- **ORF** : Open Reading Frame.
- **WHO** : World Health Organisation.

Résumé

Poussées par la sélection naturelle, la dérive génétique aléatoire, l'édition de gènes et les réponses immunitaires de l'hôte, les virus évoluent constamment à travers des mutations, qui créent de la diversité génétique et génèrent de nouveaux variants. Pour bien comprendre comment la mutation affectera l'infectiosité, la transmission et la virulence du SARS-CoV-2, il sera d'une grande importance d'étudier les mutations sur le SARS-CoV-2, en particulier ses gènes les plus variables dont la protéine S et son domaine de liaison des récepteurs. Par conséquent, dans ce travail, nous nous concentrons principalement sur ces mutations et les mutations de la protéine de la pointe S.

Pour réaliser une étude comparative entre les variants et les mutations qui les ont touchées à travers plusieurs pays au monde, nous avons traité plusieurs séquences génomiques qui ont été rapidement accumulées à la GenBank depuis janvier 2020, séquencées et publiées par plusieurs laboratoires au monde. Ces séquences nous ont permis de construire des arbres phylogénétiques en utilisant un outil de bio-informatique très puissant, amélioré dernièrement pour étudier le Covid qui est MEGA version 11. La distribution des séquences mutées de différentes souches à travers le monde dans les arbres phylogénétiques montre que la majorité des mutations pour chaque variant se retrouvent dans la protéine S, ainsi que tous les variants portant la mutation D614G et la mutation N501Y qui modifient légèrement l'extrémité de la protéine S, et augmentent l'affinité entre cette protéine et le récepteur de conversion de l'angiotensine 2 des cellules de l'hôte. Mais en plus de la mutation D614G le variant delta combine entre les deux mutations L452R et E484K qui sont responsables de l'échappement de la réponse immunitaire de l'hôte ; c'est ce qui rend le variant delta plus dangereux et plus transmissible que les autres variants.

Mots clés : SARS-CoV-2, phylogénétique, bioinformatique, gènes, mutations, variants.

ملخص

مدفوعاً بالانتقاء الطبيعي، الانحراف الجيني العشوائي، تحرير الجينات والاستجابات المناعية للمضيف، تتطور الفيروسات باستمرار من خلال الطفرات التي تخلق التنوع الجيني وتولد متغيرات جديدة. لفهم الكيفية التي ستؤثر بها الطفرة على العدوى ، سيكون مهما دراسة مكان الطفرات على SARS- CoV-2 ، وخاصة الجينات الأكثر تغيراً. بما في ذلك البروتين S ومجال ربط مستقبلاته. لذلك ، في هذا العمل ، نركز بشكل أساسي على هذه الطفرات والطفرات في البروتين S. لإجراء دراسة مقارنة بين المتغيرات والطفرات التي تتأثر بها عبر العديد من البلدان في العالم ، قمنا بمعالجة العديد من التسلسلات الجينية التي تراكمت بسرعة في GenBank منذ يناير 2020 ، تم تسلسلها ونشرها من قبل العديد من المختبرات في العالم. سمحت لنا هذه التسلسلات ببناء أشجار النشوء والتطور باستخدام أداة معلوماتية حيوية قوية جداً ، تم تحسينها مؤخراً لدراسة Covid وهو الإصدار 11 من MEGA. توزيع التسلسلات المتحولة من سلالات مختلفة حول العالم في أشجار النشوء والتطور يظهر أن غالبية الطفرات لكل منها تم العثور على متغير في بروتين S ، وكذلك جميع المتغيرات التي تحمل طفرات D614G و N501Y التي تعدل بشكل طفيف نهاية بروتين S ، وتزيد من التقارب بين هذا البروتين ومستقبلات مستقبلات تحويل الأنجيوتنسين 2 للخلايا المضيئة. ولكن بالإضافة إلى طفرة D614G، فإن متغير دلتا يجمع بين الطفرتين L452R و E484K وهما المسؤولتان عن هروب الاستجابة المناعية للمضيف ؛ هذا ما يجعل متغير دلتا أكثر خطورة وأعلى قابلية للانتقال من المتغيرات الأخرى.

الكلمات المفتاحية : SARS- CoV-2، علم الوراثة ، المعلوماتية الحيوية ، الجينات ، الطفرات ، المتغيرات.

Summary

Driven by natural selection, random genetic drift, gene editing and host immune responses, viruses constantly evolve through mutations, which create genetic diversity and generate new variants. To fully understand how the mutation will affect the infectivity, transmission and virulence of SARS-CoV-2, it will be of great importance to study mutations on SARS-CoV-2, especially its most variable genes. Including protein S and its receptor binding domain. Therefore, in this work, we mainly focus on these mutations and mutations of protein S. To carry out a comparative study between the variants and the mutations that have affected them across several countries around the world, we have processed several genomic sequences that have been rapidly accumulated at GenBank since January 2020, sequenced and published by several laboratories around the world. These sequences allowed us to build phylogenetic trees using a very powerful bioinformatics tool, recently improved to study Covid which is MEGA version 11. The distribution of mutated sequences from different strains around the world in the phylogenetic trees shows that the majority of the mutations for each variant are found in the S protein, as well as all the variants carrying the D614G mutation and the N501Y mutation which slightly modify the end of the S protein, and increase the affinity between this protein and the receptor of host cell Angiotensin 2 conversion. But in addition to the D614G mutation the delta variant combines between the two mutations L452R and E484K which are responsible for the escape of the host immune response; this is what makes the delta variant more dangerous and more transmissible than other variants.

Keywords: SARS-CoV-2, phylogenetics, bioinformatics, genes, mutations, variants.

Table de matières

Chapitre I : Recherches bibliographiques

SARS CoV-2.....	1
Historique	1
Présentation	2
Taxonomie et classification	2
Origine du SARS-CoV-2.....	4
Structure et génome du SARS-CoV-2.....	4
Structure du virus.....	5
Structure du génome	5
Modes de transmission	8
Mutations et variants du SARS-CoV-2	9
Définitions des variants du SARS-CoV-2	9
Variants préoccupants « variant of concern »	11
Impact des variants sur la pathogénie du SARS-CoV-2.....	14
Généralités sur la bio-informatique.....	15
Rôle de la bio-informatique dans l'étude de la physiopathologie du SARS-CoV-2 et variants....	15
Banques des données biologiques	15
Différentes étapes et logiciels d'une étude phylogénétique	17
Etude phylogénétique.....	19
Définition de la phylogénie	19
Phylogénie moléculaire	19
Marqueurs d'ADN nucléaire	19
Marqueurs d'ADN mitochondrial	19
Représentation phylogénétique	20
Arbre phylogénétique	20
Caractéristiques générales des arbres	21
Différentes représentations graphiques	22
Reconstruction d'arbre phylogénétique	23

Chapitre II : Matériel et méthodes

Extraction des données	24
Présentation de base de données NCBI	24
Etapes de la recherche sur NCBI.....	24
Traitement des séquences de la GenBank.....	29
Utilisation de FaBox.....	30
Recherche des similarités par BLAST	32
Présentation de MEGA version 11	35
Alignement des séquences.....	36
Construction des arbres phylogénétiques	39

Chapitre III : Résultats et Discussion

III.1.Résultats	42
GenBank.....	42
Analyse phylogénétique.....	45
Construction des arbres phylogénétiques par MEGA 11	46
Arbre phylogénétique du gène de pointe S	48
Arbre phylogénétique du gène ORF1ab	54
Arbre phylogénétique du gène N	58
Arbre phylogénétique du gène ORF1ab	62
Discussion.....	65
Comparaison des bases azotées des gènes plus variables pour le SARS-CoV-2	65
Comparaison entre les séquences des variants de différents pays dans le monde	65
Les différentes mutations du Coronavirus	67
Vaccin et mutations.....	74
III.3.Covid-19 en Algérie	75

Conclusion

Références bibliographiques

Biblio-web

INTRODUCTION

Introduction

Le COVID-19 est une maladie causée par le SARS-CoV-2 et constitue une menace majeure pour la santé publique au niveau mondial en raison des taux élevés d'infection et de mortalité. La morbidité et la mortalité continuent d'augmenter en raison de l'absence d'un médicament spécifique qui empêche la progression de la maladie COVID-19, malgré la présence de certains vaccins au cours du premier semestre 2020 en particulier de mars à mai, le nombre des personnes atteintes par le COVID-19 a explosé, tandis que les pays déploient de grands efforts pour s'adapter à l'émergence de nouveau virus pandémique (**Wu et al., 2020**).

Le premier cas de SARS-CoV-2 a été signalé le 30 décembre 2019 dans la province de Wuhan en Chine et a été déclaré par la suite pandémie mondiale (**Hang et al., 2020**) dont les principaux symptômes sont la toux, la fièvre et les difficultés respiratoires, ces symptômes peuvent évoluer vers une détresse respiratoire et une insuffisance organique. En outre, la maladie peut provoquer une inflammation systémique et sa transmission se fait par l'infection par des gouttelettes, mais d'autres modes tels que la transmission aéroportée et la transmission oro-fécale sont également spéculés (**Wang et al., 2020**). L'Organisation mondiale de la santé a annoncé une urgence de santé publique dans le monde entier pour cette pandémie. Selon cette dernière, environ 205338159 cas de Covid ont été confirmés dans le monde et plus de 433094 de décès sont recensés depuis le 31 décembre 2019. L'Algérie compte désormais 224 979 cas confirmés, 154 131 guérisons et 6 393 décès.

Le SARS-CoV-2 est un virus enveloppé avec un génome à ARN simple brin de polarité positive d'environ 29,8 à 29,9 kb avec 10 cadres de lecture ouverts (ORF) codant pour environ 30 protéines. L'ORF1ab code pour 16 protéines non-structurales, dont la RdRp, les autres ORF codent entre autre, pour quatre protéines structurales ; la glycoprotéine S, qui est l'épitope des anticorps neutralisants et le site d'attache au récepteur ACE, la protéine N (nucléoprotéine), la protéine d'enveloppe E et la matrice M. Les gènes codant pour la RdRp et les protéines structurales sont les cibles recherchées par les tests PCR. Etant un virus à ARN, les mutations sont habituelles et peuvent engendrer des variants d'intérêt tel que les variants Alpha, Bêta, Gamma, Delta ou encore fin Novembre 2021 Omicron (**Xu et al., 2021**).

L'émergence de nouveaux variants rapidement transmissibles du SARS-CoV-2 menacent de prolonger cette pandémie et d'avoir des conséquences sanitaires et économiques dévastatrices, les principaux variants des préoccupations (VOC) ont des mutations communes dans les

Protéines de pointe du SARS-CoV-2, principalement sur l'unité S1, ce qui entraîne un taux de transmissibilité plus élevé et affecte la virulence virale et le résultat clinique. Les mutations de la protéine spike et d'autres mutations de protéines dans les VOC peuvent permettre d'échapper aux vaccinations approuvées dans certains cas (**Cheen et al., 2021**).

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**). Pour la réalisation de l'arbre phylogénétique, il existe cependant quelques limites à l'utilisation de l'outil bio-informatique pour la reconstruction de l'arbre phylogénétique. Cette reconstruction se base sur le fait que certains caractères ont été hérités d'une espèce à l'autre, tandis que d'autres se sont transformés au cours du temps, laissant ainsi des traces de l'évolution (**Darlu et al., 1993**).

Le présent travail a pour objectif d'étudier les gènes les plus variables de coronavirus qui sont: ORF1ab, S, ORF3a et N, les variants et les mutations détectés dans leurs séquences génomiques. Nous portons notre attention sur les mutations qui se sont produites au niveau de ces gènes car elles jouent un rôle clé dans la pénétration du virus au sein des cellules de l'hôte, pour cela nous avons fait appel à la bioinformatique afin de réaliser des analyses phylogénétiques tel que la construction des arbres phylogénétiques.

Notre travail est basé sur une étude phylogénétique commence par le chapitre I recherche bibliographique, qui se divise en trois parties :

- ✚ Présentation de notre modèle d'étude ; le genre SARS-CoV-2, classification, structure, mode de transmission, mutations, variants.
- ✚ Définir : la phylogénie, l'arbre phylogénétique et les caractéristiques générales de l'arbre qui aident à la reconstruction phylogénétique des gènes.
- ✚ Présentation des différentes banques de données et les différentes étapes d'une étude phylogénétique

Le chapitre II, fournit le matériel et les méthodes que nous utiliserons pour réaliser notre mémoire.

Le chapitre III, rapporte les résultats obtenus ainsi qu'une discussion à la lumière des données bibliographiques.

Enfin, on terminera ce mémoire par une conclusion générale où nous dégagerons les principaux résultats ainsi que les perspectives ouvertes par ce travail.

RECHERCHES
BIBLIOGRAPHIQUES

Recherches bibliographiques

1.1. SARS Covid 2

1.1.1. Historique

En 1930, les chercheurs américains **Schalk** et **Hawn** ont signalé une nouvelle maladie respiratoire chez les poussins, se caractérisant par une détresse respiratoire aiguë. Le virus responsable est appelé : virus de la bronchite infectieuse aviaire (Infectious Bronchitis Virus, IBV).

En 1965, les chercheurs britanniques **David Tyrrell** et **Malcolm Bynoe** caractérisent à partir d'écouvillonnage respiratoire d'un jeune garçon présentant un rhume typique une souche virale, appelée B 814.

En 1966, les chercheurs **Hamre** et **Procknow** de l'Université de Chicago ont rapporté qu'ils avaient cultivés dans des cultures de cellules rénales embryonnaires humaines inoculées avec des échantillons des voies respiratoires d'un patient atteint de rhume, la souche virale, appelée 229 E, dont le matériel génétique est composé d'ARN.

En 1967, **Kenneth McIntosh** des National Institutes of Health (NIH, Bethesda, Maryland) et ses collaborateurs ont découvert une nouvelle souche de virus appelée OC43, dont la forme était très similaire au virus de la bronchite infectieuse aviaire. Ce virus a été isolé à partir d'explants de tubes humains stockés en culture d'organes, novembre 1968, le magazine "Nature" a rapporté qu'un groupe de virologues avait déterminé le nom "coronavirus" pour un nouveau groupe de virus sur la base de critères morphologiques complets.

Entre novembre 2002 et juillet 2003, un agent infectieux a causé une épidémie de pneumonie atypique dans la province de Guangdong, Sud de la Chine se caractérisant généralement par une fièvre élevée et des symptômes respiratoires légers, mais évoluant rapidement vers une pneumonie en quelques jours. L'agent responsable du Syndrome Respiratoire Aigu Sévère (SRAS) est nommé SARS-CoV ayant infecté 8437 individus et causé 813 décès dans le monde représentant ainsi la première pandémie bien documentée de ce siècle.

Une épidémie de pneumonies de cause inconnue a été découverte et décrite à Wuhan, dans la province du Hubei, en Chine, en décembre 2019. La découverte d'un nouveau coronavirus a été déclarée urgence de santé publique de portée internationale par l'Organisation mondiale de la santé (OMS) le 9 janvier 2020. Appelé pour la première fois 2019-nCoV puis SARS-CoV-

Recherches bibliographiques

2, ce virus est différent du virus SARS-CoV et il est également différent du virus MERS-CoV. La pandémie COVID-19 a été déclarée par l'OMS le 11 mars 2020 (**Meriadeg et al., 2020**).

I.1.2. Présentation

Les coronavirus (CoV) appartiennent au genre Coronavirus, à la famille des Coronaviridae et à l'ordre des Nidovirales. Ils sont enveloppés et ont un acide ribonucléique non segmenté, simple brin, à sens positif (ssRNA +) comme matériel nucléaire. En microscopie électronique, ces virus présentent un aspect caractéristique qui ressemble à une couronne (corona en latin signifie couronne) en raison de la présence de projections protéiques de surface en forme de massue. Les CoV sont pléomorphes, mesurent entre 80 et 160 nm de longueur et ont un petit génome mesurant 27 à 32 kilo-octets (Ko) avec une stratégie de réplication unique (**Mahendra et al., 2020**). Selon des analyses génomiques comparatives combinées à la synthèse d'arbres évolutifs, la souche initiale du virus SARS-CoV-2 appartient à la lignée b du genre beta-CoV. Il partage également une homologie phylogénétique significative avec le type BatCoVraTG13 (**Wu et al., 2020**).

I.1.3. Taxonomie et classification

I.1.3.1. Taxonomie

Selon le Comité international de taxonomie des virus, les CoV sont classés dans l'ordre des Nidovirales, une famille de Coronaviridae et une sous-famille de Coronavirinae (**Fehr et al., 2015**). Sur la base de preuves sérologiques antérieures et génomiques récentes, la famille des Coronaviridae comprend deux sous-familles : la sous-famille Orthocoronavirinae et la sous-famille Torovirinae citées dans la figure 1 (**Ashour et al., 2020**). La sous-famille des Orthocoronavirinae comprend quatre genres : Alpha coronavirus, Beta coronavirus, Gamma coronavirus et Delta coronavirus (**Cui et al., 2019**). Les alphas et bêtacoronavirus infectent les mammifères, les gammacoronavirus infectent les espèces aviaires et les deltacoronavirus infectent à la fois les espèces mammifères et aviaires. Les alphacoronavirus représentatifs comprennent le coronavirus humain NL63 (HCoV-NL63), le coronavirus de la gastro-entérite transmissible porcine (TGEV), le PEDV et le coronavirus respiratoire porcin (PRCV). Les bêtacoronavirus représentatifs comprennent le SARS-CoV, le MERS-CoV, le coronavirus de chauve-souris HKU4, le coronavirus de l'hépatite de la souris (MHV), le coronavirus bovin (BCoV) et le coronavirus humain OC43.

Recherches bibliographiques

Les gammas et deltacoronavirus représentatifs la bronchite infectieuse aviaire (IBV) et le deltacoronavirus porcin (PdCV) (Hozhabri *et al.*, 2020).

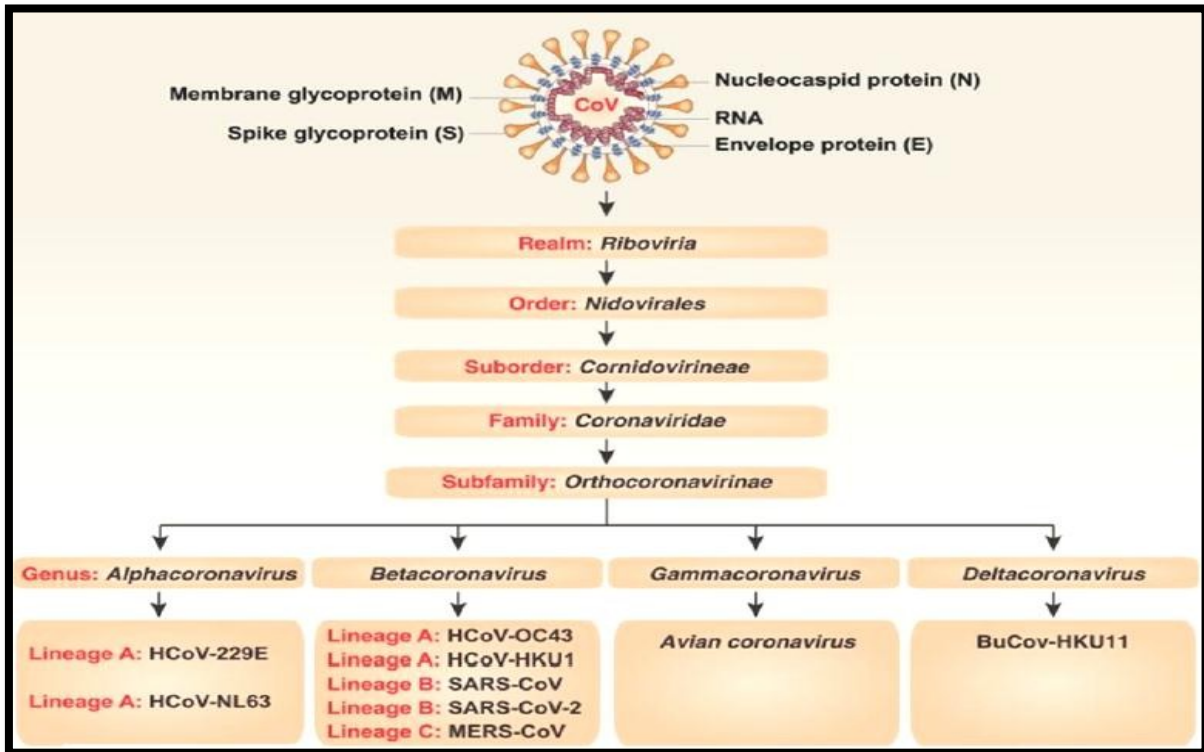


Figure 1 : Schéma représentatif de la classification de Coronaviridae (Tang *et al.*, 2020).

I.1.3.2. Origine du SARS-CoV-2

Le génome du SARS-CoV-2 présente 79,6 % et près de 50 % de similitude avec les coronavirus de type SARS-CoV et MERS-CoV respectivement. Le SARS-CoV-2 présente 96 % d'homologie au niveau du génome entier avec le coronavirus de la chauve-souris. Il existe sept domaines de réplication conservés dans le gène ORF1ab du SARS-CoV-2 qui partagent une identité de séquence de 94,4 % avec le SARS-CoV. En outre, une courte région RdRp provenant d'un coronavirus de chauve-souris appelé BatCoV RaTG13 a démontré une identité de séquence élevée avec le SARS-CoV-2. Le séquençage complet de cet ARN a donné lieu à une concordance de séquence génomique complète de 96,2 %. La protéine de liaison aux récepteurs spike S du SARS-CoV-2, qui présentait une grande diversité dans d'autres CoV était identique à 93,1 % au gène S de RaTG13, à l'exception de trois courtes insertions dans l'extrémité N-terminale et quatre changements de résidus clés sur cinq dans le motif de liaison aux récepteurs. En outre, l'étude de la région codante du génome du SARS-CoV-2 a montré que les nucléotides et les acides aminés de cette région présentent une ressemblance de 92,67

Recherches bibliographiques

% et 96,92 % au niveau des nucléotides et 97,82 % et 98,67 % au niveau des acides aminés avec le génome du CoV du pangolin et de la chauve-souris. Ces données phylogénétiques soutiennent la théorie de l'origine chauve-souris du SARS-CoV-2 et que le SARS-CoV-2 pourrait avoir évolué naturellement à partir du coronavirus de chauve-souris RaTG13. Cependant, des recherches supplémentaires sont encore nécessaires pour clarifier la présence d'un hôte intermédiaire qui a favorisé la transmission du virus, car certains éléments indiquent que le virus n'a pas été transmis de la chauve-souris à l'homme (**Meriadeg et al., 2020**).

I.1.4. Structure et génome du SARS-CoV-2

I.1.4.1. Structure du virus

La figure 2 représente un schéma de la particule virale du SARS-CoV-2 qui est sphérique avec une taille moyenne de 120 nm. Un aspect en couronne des spicules de l'enveloppe virale en microscopie électronique d'où le nom de Coronavirus. La nucléocapside est constituée de la protéine de capsid (N) complexée à l'ARN virale et protégé par une enveloppe phospholipidique portant à sa surface la glycoprotéine de surface : la protéine Spike (S), protéine d'enveloppe (E) ainsi que la matrice (M). L'entrée des coronavirus dans les cellules hôtes est médiée par la glycoprotéine Spike (protéine S). Les glycoprotéines Spike transmembranaires forment des homotrimères qui dépassent de la surface virale. La glycoprotéine Spike est essentielle pour l'entrée des coronavirus et constitue donc une cible antivirale. La figure 3 montre que la protéine S est composée de deux sous-unités fonctionnelles, dont les sous-unités S1 et S2 ; la sous-unité S1 est constituée d'un domaine N-terminal (NTD) et un domaine de liaison au récepteur (RBD) sa fonction est de se lier au récepteur de la cellule hôte, la sous-unité S2 contient le peptide de fusion (FP), la répétition heptadique 1 (HR1), une hélice centrale (CH), un domaine connecteur (CD), les glycanes N-liés sont essentielles pour la neutralisation des anticorps et la forme des trimères de protéines. Dans l'ensemble, la structure de la protéine S du SARS-CoV-2 ressemble à celle de la protéine S du SARS-CoV. Dans la conformation de préfusion, les sous-unités S1 et S2 restent liées de manière non covalente.

Différents types de coronavirus utilisent des domaines spéciaux dans la sous-unité S1 pour reconnaître différents récepteurs d'entrée. Dans le cas du SARS-CoV et SARS-CoV-2, la protéine S présente deux formes de structure, dont l'état fermé et l'état ouvert. À l'état fermé, les trois motifs de reconnaissance ne dépassent pas de l'interface formée par trois protomères de la protéine S (**Meriadeg et al., 2020**).

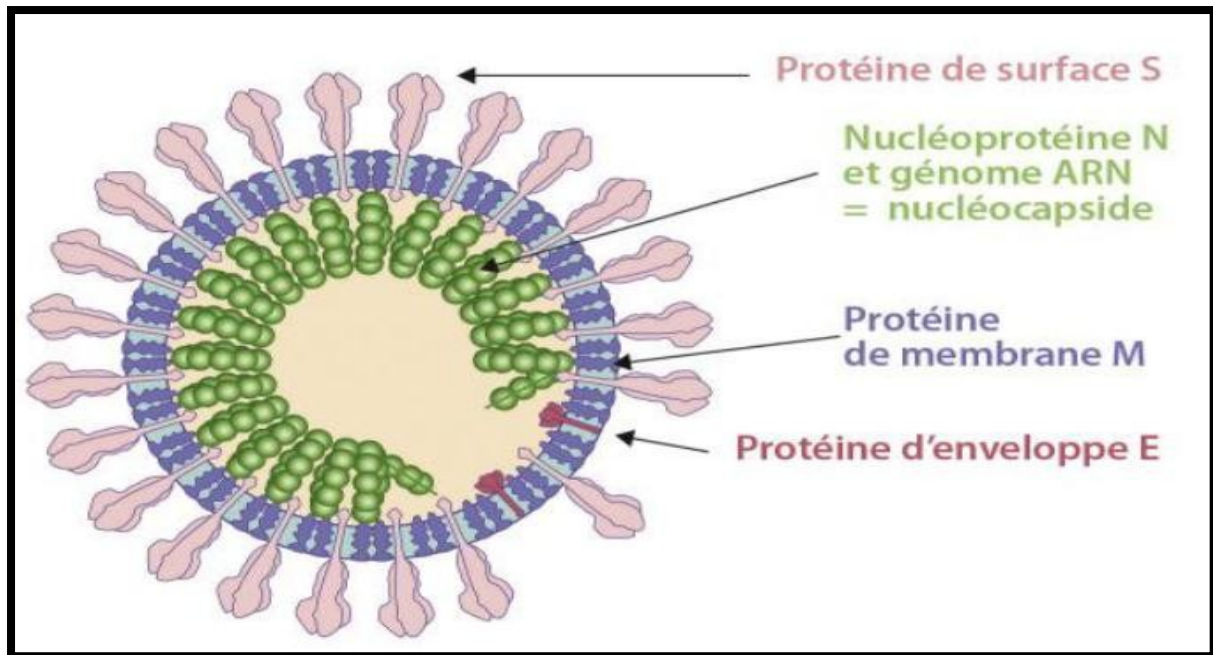


Figure 2: Structure schématisée du SARS-CoV-2 (Gautheret, 2020).

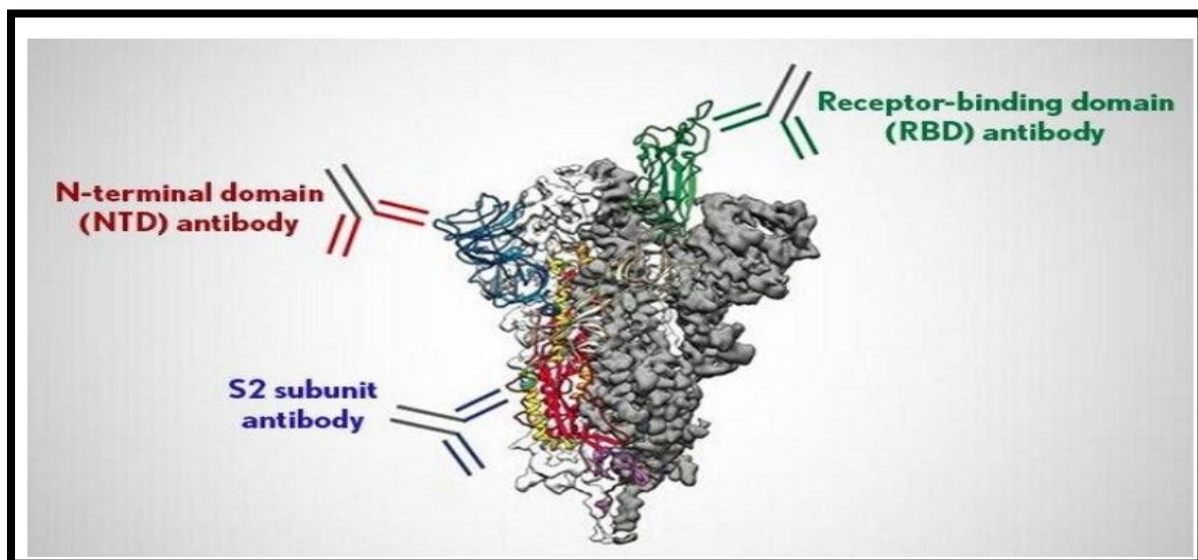


Figure 3 : Le domaine RBD (Receptor Binding Domain) est la région de la protéine spike entrant en contact avec le récepteur cellulaire ACE2 qui sert de porte d'entrée au SARS-CoV-2 dans les cellules qu'il infecte (Gazzali et al., 2020).

I.1.4.2. Structure du génome

Le SARS-CoV-2 est un virus à ARN simple brin de polarité positive d'environ 30 Kb et sa structure génomique suit les caractéristiques des gènes de coronavirus connus cités dans le tableau 1. Comme montre la figure 4, la polyprotéine ORF1ab couvre les deux tiers du génome virale est clivée en de nombreuses protéines non structurales (nsp1 à nsp16). La troisième partie du génome du SARS-CoV-2 code pour les principales protéines structurales

Recherches bibliographiques

;Gp spike (S), enveloppe (E), nucléocapside (N), et membrane (M). En outre, six ORF, à savoir ORF3a, ORF6, ORF7a, ORF7b, ORF8 et ORF10, n'ont pas de fonction connue. Nsp1 assure la médiation du traitement de l'ARN et la réplication. Nsp2 module la voie de signalisation de la survie de la cellule hôte. Nsp3 est censé séparer la protéine traduite. Nsp4 contient le domaine transmembranaire 2 (TM2) et assure le réarrangement de la membrane. Nsp5 participe au processus de la polyprotéine pendant la réplication (Meriadeg *et al.*, 2020).

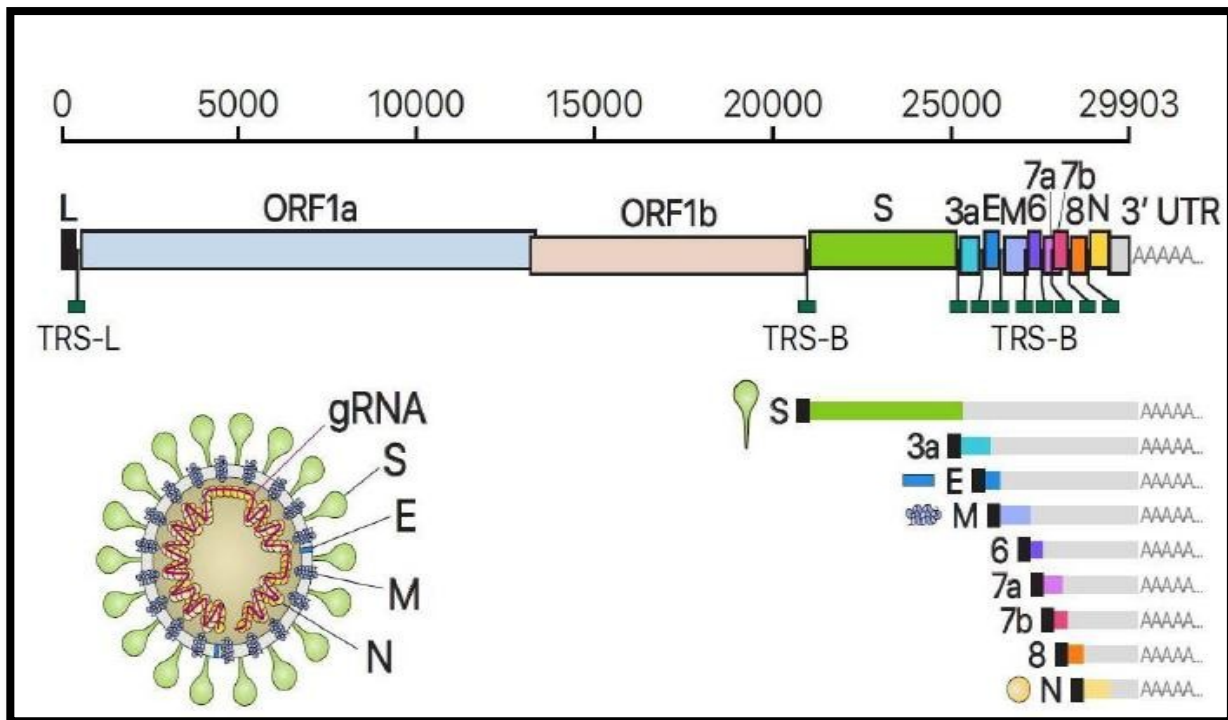


Figure 4: Présentation schématique de l'organisation du génome du SARS-CoV-2, des ARNm sous-génomiques canoniques et de la structure du virion (Meriadeg *et al.*, 2020).

Nsp 6 est un domaine transmembranaire présumé domaine. La présence de nsp7 et nsp8 augmente de manière significative l'association de nsp12 et de l'ARN modèle-primaire. Nsp9 fonctionne comme une protéine de liaison à l'ARNss. Nsp10 est critique pour la méthylation de la coiffe des ARNm viraux. Nsp12 contient l'ARN polymérase ARN dépendante (RdRp), qui est un élément critique de la réplication/transcription des coronavirus. Nsp13 se lie à l'ATP et le domaine de liaison au zinc de nsp13 participe au processus de réplication et de transcription. Nsp14 est un domaine exoribonucléase de relecture. Nsp15 possède une activité endoribonucléase dépendante du Mn (2+). Nsp16 est une 2'-Oribase méthyltransférase (Meriadeg *et al.*, 2020).

Recherches bibliographiques

Numéro	Gènes	Locus	Taille (aa)	Protéines
1	ORF1ab	BCB15089.1/ BCB97900.1	7096	Polyprotéine ORF1ab
1	ORF1a	YP_009725295.1	4405	Polyprotéine ORF1a
2	ORF2 (S)	BCA87361.1	1273	Péplomère (protéine S)
3	ORF3a	BCA87362.1	275	Protéine ORF3a
4	ORF4 (E)	BCA87363.1	75	Protéine d'enveloppe (protéine E)
5	ORF5 (M)	BCA87364.1	222	Protéine de membrane (protéine M)
6	ORF6	BCA87365.1	61	Protéine ORF6
7	ORF7a	BCA87366.1	121	Protéine ORF7a
8	ORF7b	BCB15096.1	43	Protéine ORF7b
9	ORF8	BCA87367.1	121	Protéine ORF8
10	ORF9 (N)	BCA87368.1	419	Phosphoprotéine de la nucléocapside (protéine N)
11	ORF10	BCA87369.1	38	Protéine ORF10

Tableau 1: Caractéristiques des gènes exprimés par le SARS-CoV-2 (Astrid *et al.*, 2020).

I.1.5. Modes de transmission

Comprendre le mode, l'heure et le type de transmission du SARS-CoV-2 est essentiel pour le développement des mesures efficaces de la santé publique et de la prévention des infections visant à briser la chaîne de transmission. Le SARS-CoV-2 se transmet principalement d'une personne à l'autre par contact direct, indirect ou étroit par les gouttelettes respiratoires, la salive ou par les sécrétions respiratoires, soit en étant inhalé, soit en se déposant sur les surfaces muqueuses, y compris les aérosols produits lors de la toux et de la parole.

Le virus peut être transmis indirectement en touchant une surface ou un objet préalablement contaminé par le virus, puis en se touchant le visage, les yeux ou la bouche.

A noter que la transmission par contact avec les objets contaminés est considérée comme possible, bien qu'ils n'aient pas isolé de virus cultivable en situation réelle. La propagation du

Recherches bibliographiques

SARS-CoV-2 se produit principalement chez les personnes symptomatiques, mais elle peut également se produire avant qu'elles ne présentent des symptômes, lorsqu'elles sont en contact avec d'autres personnes pendant des périodes prolongées. Si une personne qui ne développe jamais de symptômes peut également transmettre le virus à d'autres, on ne sait pas dans quelle mesure cela se produit, des recherches supplémentaires sont nécessaires dans ce domaine (Astrid et al., 2020).

Modalités du contact	Faible densité de personnes			Forte densité de personnes		
	En extérieur bien ventilé	En intérieur bien ventilé	Mal ventilé	En extérieur bien ventilé	En intérieur bien ventilé	Mal ventilé
Contact bref avec un masque						
Se taire						
Parler						
Crier						
Contact prolongé avec un masque						
Se taire						
Parler						
Crier						
Contact bref sans masque						
Se taire						
Parler						
Crier						
Contact prolongé sans masque						
Se taire						
Parler						
Crier						

Figure 5 : Risque de transmission du SARS-CoV-2 dans différentes situations. En vert : risque faible. En orange : risque modéré. En rouge : risque fort (Jones et al., 2020).

I.1.6. Mutations et variants du SARS-COV-2

I.1.6.1. Définitions des variants du SARS-COV-2

Les virus évoluent dans le temps par des mutations et tout particulièrement lorsqu'il y a beaucoup d'infections dans une population.

Les mécanismes d'évolution des virus impliquent des mutations (changements du code génétique) et des recombinaisons. Ces mutations arrivent au hasard et se produisent pendant la réplication virale à l'intérieur des cellules de l'hôte. Si le virus est porteur d'une ou plusieurs mutations par rapport au virus initial alors il est désigné sous le nom de « variants ». Cependant si le comportement de ce variant est différent (plus transmissible, plus pathogène, infectant d'autres types cellulaires) on parlera alors de « souche » virale.

Les mutations les plus fréquentes sont :

- Des substitutions : une base nucléique est remplacée par une autre ce qui peut entraîner une modification de l'acide aminé correspondant.

Recherches bibliographiques

- Des délétions : une ou plusieurs bases nucléiques sont éliminées et par conséquent la perte de l'acide aminé correspondant également.
- Des insertions : des triplets de bases sont introduits et donc de nouveaux acides aminés (Corum et al., 2021).

I.1.6.2. Variants du SARS-COV-2

Comme les autres virus à ARN le SARS-CoV-2 est sujet à des évolutions génétiques qui peuvent donner naissance à de multiples variants pouvant présenter des caractéristiques différentes de celles de la souche initiale. Certaines modifications peuvent être sans conséquence alors que d'autres peuvent être avantageuses. Les mutations se produisent sur toute la longueur du génome du virus bien qu'environ 50 % se produisent au niveau du gène de la protéine Spike (gène S), les mutations définissent chaque nouvelle lignée du SARS-COV-2. Lorsque les mutations se produisent dans la protéine S et en particulier dans la région RBD, elles peuvent affecter la liaison avec les récepteurs de l'hôte ou avec les anticorps (Konings et al., 2021).

Avec l'émergence rapide de plusieurs variants, l'OMS a établi un système de classification pour distinguer les variants émergents du SARS-CoV-2 :

Le variant G614, en ce moment cette mutation se trouve dans presque toutes les séquences du SARS-COV-2 dans le monde, c'est la première mutation observée dans le SARS-COV-2 au niveau du gène de la protéine S qui consiste en une substitution de l'acide aspartique par la glycine en position 614 améliorant la liaison entre la protéine S et le récepteur ACE2 des cellules cibles (Konings et al., 2021).

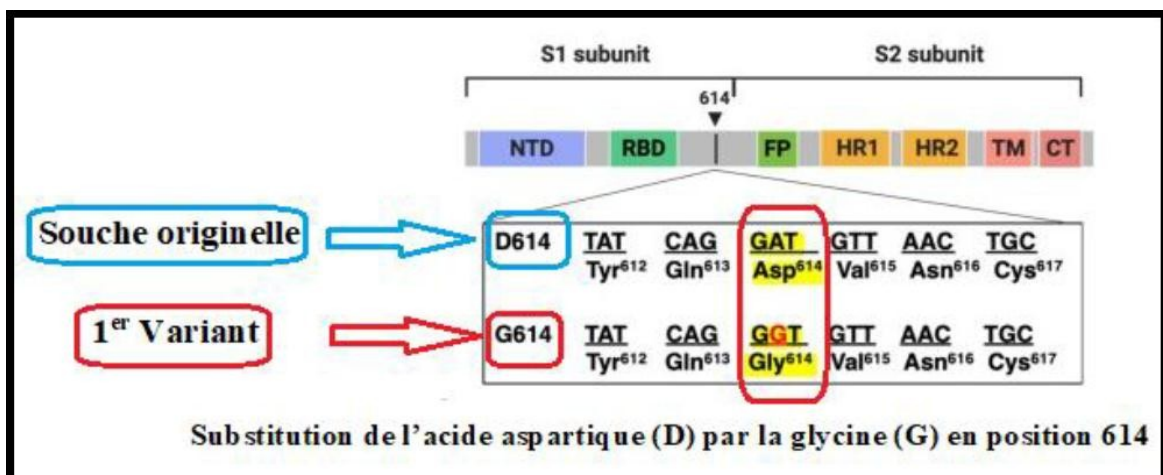


Figure 6 : La mutation D614G au niveau de la protéine S du SARS-CoV-2 (Zhang et al., 2021).

Recherches bibliographiques

I.1.6.3. Variants préoccupants « variant of concern »

Ce sont les variants pour lesquelles il existe des preuves de changements phénotypiques par rapport à la souche initiale et qui entraînent :

- Une augmentation de la transmissibilité du virus.
- Une réduction significative de la neutralisation du virus par les anticorps générés lors d'infection ou vaccination antérieure.
- Une diminution de l'efficacité des traitements ou des vaccins.
- Une augmentation de la sévérité de la maladie (augmentation des hospitalisations ou des décès).
- L'échec de détection par des tests spécifiques (**Konings et al., 2021**).

A. Variant Anglais « Alpha »

Le variant Alpha a été identifié pour la première fois en décembre 2020 au Royaume-Uni dans un test commercial fréquemment utilisé, le « TaqPath RT-PCR », qui permet par la technique RT-PCR « Reverse transcriptase PCR- polymerase chain reaction » de détecter trois cibles différentes au niveau de l'ARN viral, cependant le variant alpha comporte la délétion 69-70 ou encore désignée sous le nom $\Delta H69/V70$ car elle provoque la perte des acides aminés histidine en position 69 et valine en position 70 au niveau de la protéine S entraînant ainsi l'obtention de deux signaux au lieu de trois et donc l'absence de détection du gène S (S gene target failure ou SGTF). Le variant alpha comprend dix-sept mutations dans le génome dont neuf mutations se trouvent au niveau du gène de la protéine S (**Konings et al., 2021**).

- **S : Délétion $\Delta H69-V70$**
- **S : Délétion $\Delta Y144$** : Délétion de la tyrosine en position 144. Elle confère une résistance aux anticorps monoclonaux 4A8.
- **S : N501Y** : signifie qu'en position 501 de la chaîne des acides aminés de la protéine S une tyrosine (Y) a remplacé une asparagine (N).
- **S : A570D** : substitution de l'alanine par l'acide aspartique en position 570.
- **S : D614G**.
- **S : P681H** : substitution de la proline par l'histidine en position 687. Mutation proche du site de clivage S1/S2.
- **S : T716I** : substitution de la thréonine par l'isoleucine en position 716.
- **S : S982A** : substitution de sérine par l'alanine en position 982.

Recherches bibliographiques

- **S : D1118H** : substitution de l'acide aspartique par l'histidine (H) en position 1118 (Konings *et al.*, 2021).

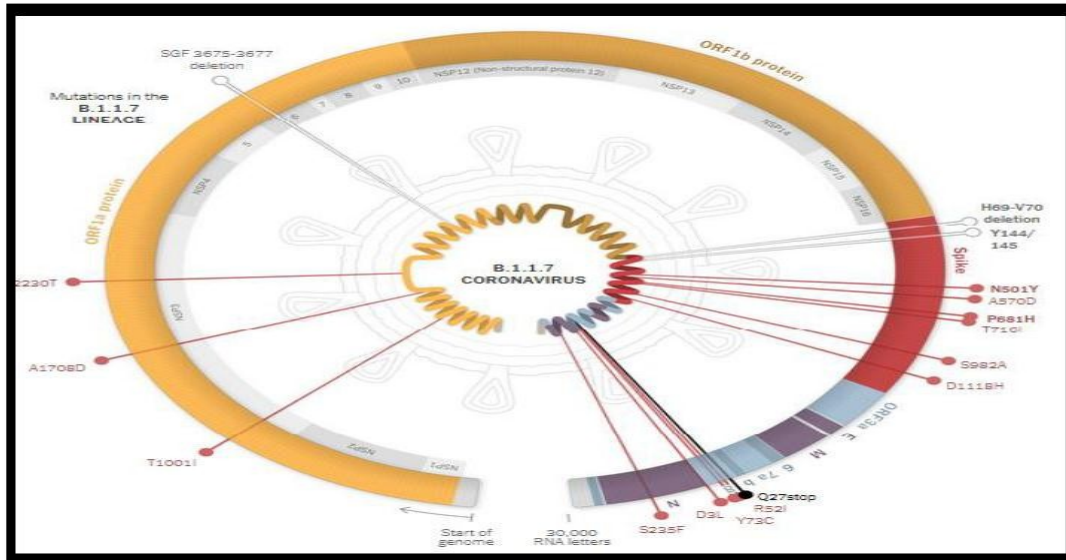


Figure 7 : Les mutations clés présentes dans le génome du variant Anglais (Segondy *et al.*, 2020).

B. Variant Sud-Africain « Béta »

Le variant Béta comprend neuf mutations (L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G et A701V) dans la protéine S dont trois mutations sont situées dans le RBD (K417N, E484K et N501Y):

- **S : L18F***: substitution de la leucine (L) par la phénylalanine (F) en position 18.
- **S : D80A** : substitution de l'acide aspartique (D) par l'alanine (A) en position 80.
- **S : D215G** : substitution de l'acide aspartique (D) par la glycine (G) en position 215.
- **S : Délétion L241/L242** : délétion de la leucine (L) en position 241/242.
- **S : Délétion A243** : Délétion de l'alanine (A) en position 243.
- **S : K417N** : substitution de la lysine (K) par l'asparagine (N) en position 417.
- **S : E484K** : substitution de l'acide glutamique (E) par la lysine (K) en position 484.
- **S : N501Y, D614G.**
- **S : A701V** : substitution de l'alanine (A) par la valine en position 701 (Larsenet *et al.*, 2020).

(*): Mutation présente que dans certaines sous souche.

C. Variant Brésilien « Gamma »

Recherches bibliographiques

Le variant Gamma héberge douze mutations dans la protéine S (L18F, T20N, P26S, D138Y, R190S, H655Y, T1027I, V1176F, K417T, E484K et N501Y) dont trois mutations (L18F, K417N, E484K) sont localisées dans le RBD similaire au variant Béta.

- S : L18F
- S : T20N : substitution de la thréonine (T) par l'asparagine (N) en position 20.
- S : P26S : substitution de la proline (P) par la sérine (S) en position 26.
- S : D138Y : substitution de l'acide aspartique (D) par la tyrosine (Y) en position 138.
- S : R190S : substitution de l'arginine (R) par la sérine en position 190.
- S : K417T, E484K, N501Y, D614G.
- S : H655Y : substitution de l'histidine par la tyrosine en position 655.
- S : T1027I : substitution de la thréonine (T) par l'isoleucine (I) en position 1027.
- S : V1176F : substitution de la valine (V) par la phénylalanine (F) en position 1176.

Le variant Brésilien semble présenter un profil de mutations similaire au variant sud-africain (K417N/T, E484K, N501Y) (Larsenet *al.*, 2020).

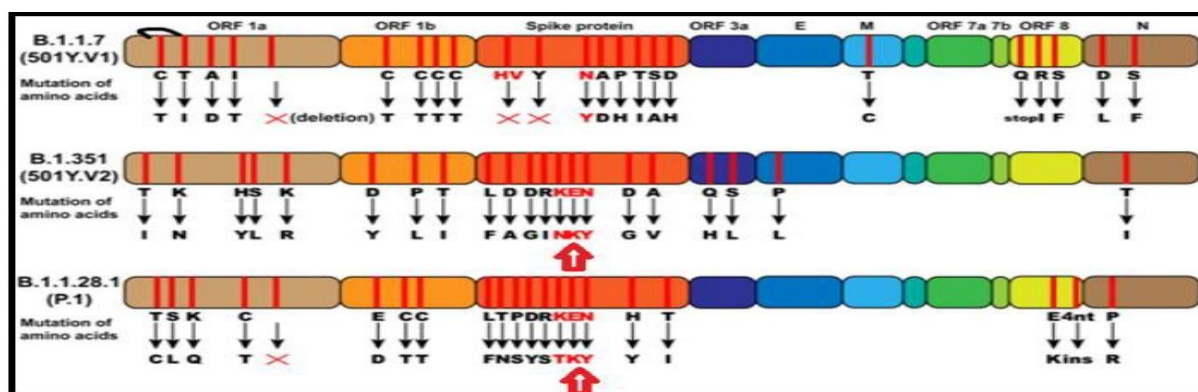


Figure 8 : Les mutations observées dans les variants « Alpha, Béta, Gamma (Zhou W, *al.*, 2020).

D. Variant Indien « Delta »

Il a été détecté pour la première fois en Inde en décembre 2020 et est porteur de plus d'une douzaine de mutations dont 6 mutations au niveau de la protéine S :

- S : E154K : substitution de l'acide glutamique (E) par la lysine (L) en position 154.
- S : L452R : substitution de la leucine (L) par l'arginine (R) en position 452.
- T478K : substitution de la thréonine (T) par la lysine (K) en position 478
- S : E484Q : substitution de l'acide glutamique (E) par la glutamine.
- S : D614G.

Recherches bibliographiques

- **S : P681R** : substitution de la proline par l'arginine en position 681.
- **S : Q1071H** : substitution de la glutamine par l'histidine en position 1071 (**Larsen et al., 2020**).

E. Variant Américain et Epsilon

Ces deux variants sont apparus pour la première fois aux Etats-Unis en juin 2020 et sont passés de 0 % à plus de 50 % des cas recensés entre le 1er septembre 2020 et le 29 janvier 2021. Actuellement ils sont considérés comme variants préoccupants aux Etats-Unis alors que l'OMS les classe comme variants à suivre.

- Le **B.1.427** comprend deux mutations **D614G, L452R**
- Le **B.1.429** comprend quatre mutations :
- **S13I** : substitution de la sérine (S) par l'isoleucine en position 13.
- **W152C** : substitution du tryptophane (W) par la cystéine en position 152.
- **L452R, D614G** (**Larsen et al., 2020**).

F. Variant Iota « B.1.526 »

Détecté pour la première fois à New-York en novembre 2020, il se présente sous deux sous souches : la première comprend la mutation E484K dans la protéine S. La deuxième porte la mutation S477N au niveau de la protéine S

- **S : L5F*** : substitution de la leucine (L) par la phénylalanine (F) en position 5.
- **S : T95I** : substitution de la thréonine (T) par l'isoleucine (I) en position 95.
- **S : D253G** : substitution de l'acide aspartique (D) par la glycine (G) en position 253.
- **S : S477N*** : substitution de sérine (S) par l'asparagine (N) en position 477.
- **S : E484K*, D614G**
- **S : A701V*** : substitution de l'alanine (A) par la valine (V) en position 701 (**Larsen et al., 2020**).

G. Variant Nigérien « Eta »

Initialement détecté au Nigéria et aux Etats-Unis en décembre 2020, il est porteur de la mutation E484K et de deux délétions (Δ H69/V70, Δ 144) dans la protéine Spike similaire à celles retrouvées chez le variant Alpha. Il comprend les mutations suivantes.

- **S : Q52R** : substitution de glutamine (G) par l'arginine (R) en position 52.
- **S : A67V** : substitution de l'alanine (A) par la valine (V) en position 67.
- **S : E484K**

Recherches bibliographiques

- **S : Délétion H69/V70**
- **S : Délétion Y144, D614G**
- **S : Q677H** : substitution de la glutamine (Q) par l'histidine (H) en position 677.
- **S : F888L** : substitution de phénylalanine (F) par la leucine (L) en position 888 (**Larsen et al., 2020**).

I.1.6.4. Impact des variants sur la pathogénie du SARS-COV-2

Les variants du SARS-COV-2 n'entraînent pas pour la plupart des symptômes plus graves, cependant ils ont en commun ; une transmissibilité plus élevée qui fait qu'ils deviennent rapidement dominants au sein des populations, une accumulation d'un nombre élevé de mutations dans la protéine S. Il existe également des signes d'une virulence accrue pour certains variants (variant Alpha et Delta), d'une fréquence de réinfection et d'une résistance accrue à l'action des anticorps monoclonaux et polyclonaux provenant des sérums de convalescence et des vaccins, le tableau ci-dessous résume les principales caractéristiques virologiques et cliniques de quelques variants du SARS-COV-2 (**Gómez et al., 2021**).

Variants	Alpha	Béta	Gamma	Delta	Eta
Pays	Royaume-Uni	Sud d'Afrique	Brésil/Japon	Inde	Nigéria
Détecté en Algérie	Détecté	Non détecté	Non détecté	Détecté	Détecté
Transmission	29% plus transmissible	25% plus transmissible	38 % Plus transmissible	55-97% plus transmissible	plus transmissible
Symptômes	Toux, maux de gorge, fatigue et myalgie plus fréquentes. Perte d'odorat est moins fréquente.	Similaire à ceux du variant Alpha	Similaire à ceux du SARS-COV-2	Toux persistante, maux de gorge, fièvre, écoulement nasal, douleurs musculaire , diarrhée	Similaire à ceux du SARS-COV-2
Sévérité	Augmentation probable du risque de décès : +30-70%	Possible augmentation du risque de décès : +20%	Taux de mortalité plus élevé	Taux de mortalité plus élevé	Pas de donnée disponible

Tableau 2 : Caractéristiques Cliniques et virologiques des principaux variants du SARS-CoV (**Gómez et al., 2021**).

A partir de ce tableau ils ont pu constater que le SARS-COV-2 a évolué pendant ces 20 mois de pandémie et continue encore d'évoluer à l'heure actuelle par des mutations qui entraînent

Recherches bibliographiques

L'apparition de nouveaux variants. La détermination des variants et l'analyse des mutations ainsi que l'étude phylogénétique nécessitent une discipline très importante dans le monde scientifique ; la bio-informatique qui est une discipline visant le traitement automatique de l'information biologique et dont l'alignement multiple des séquences constitue une tâche fondamentale pour plusieurs applications en bio-informatique (**Gómez et al., 2021**).

I.2. Généralité sur la bio-informatique

La bio-informatique est un domaine interdisciplinaire qui associe l'informatique et les bio-statistiques aux sciences biologiques et biomédicales telles que la biochimie, la biologie cellulaire, la biologie du développement, la génétique, la génomique et la physiologie. La bio-informatique est devenue une discipline importante peu après le développement des technologies de séquençage de l'ADN à haut débit dans les années 1970, l'objectif de la bio-informatique est de faciliter la gestion, l'analyse et l'interprétation des données issues d'expériences biologiques et des études d'observation (**Ray et al., 2021**).

I.2.1. Rôle de la bio-informatique dans l'étude de la physiopathologie du SARS COV2 et variants

L'exploitation des outils et techniques de la bio-informatique a permis d'interpréter avec succès l'architecture génomique virale du SARS-COV-2 dans un délai très court.

La bio-informatique a non seulement permis de séquencer le génome du SARS-COV-2 mais aussi d'analyser correctement les erreurs de séquençage, les relations évolutives et les variations génétiques par l'application d'un algorithme d'alignement des séquences nucléiques et protéiques. Ces données sont essentielles pour le développement d'un vaccin contre le SARS-COV-2 (**Ray et al., 2021**).

I.2.2. Banques des données biologiques

Aujourd'hui les méthodes rapides de séquençages sont utilisées fréquemment et le nombre de nouvelles séquences augmente rapidement. Toutes les données issues du séquençage doivent être traitées et analysées afin d'obtenir le plus grand nombre d'informations. Il faut ainsi stocker ces séquences et toutes les informations obtenues. Pour cela, de grandes bases de données de séquences ont été mises en œuvre pour permettre un accès facile aux données. Les premières banques de données en biologie moléculaire ont traité des informations structurales sur les protéines, puis très rapidement, des séquences protéiques et nucléotidiques. Il existe différents types de bases de données biologiques : celles qui sont dites généralistes et qui stockent des séquences provenant de tous les organismes et celles dites spécialisées qui se

Recherches bibliographiques

consacrent plus particulièrement à un organisme ou à une thématique donnée (Wu et al., 2006).

I.2.2.1. Banques de données généralistes

Il existe plusieurs banques généralistes publiquement accessibles. La principale banque généraliste de séquences nucléotidiques est produite par trois partenaires : EMBL data library (Cochrane et al., 2006) en Europe, GenBank (Benson et al., 2006) aux Etats-Unis et DDBJ (Okubo et al., 2006) au Japon. La plupart des données de ces banques proviennent de soumissions effectuées par les auteurs. D'autres regroupent des séquences protéiques telles qu'UNIPROT (Wu et al., 2006), GenPept, HAMAP (Gattiker et al., 2003), etc. De la même manière que pour les banques de séquences nucléotidiques, leur organisation se base autour des annotations biologiques et biochimiques d'une part, et des séquences d'autre part. GenPept correspond à la traduction de l'ensemble des parties codantes de GenBank. La principale banque de protéines est UNIPROT. En effet, elle possède de nombreux atouts : redondance minimale, références croisées, qualité d'annotation, etc. Elle correspond à la fusion de SWISS-PROT (Wu et al., 2006), TrEMBL et PIR (Wu et al., 2006).

Les séquences contenues dans SWISS-PROT sont issues de la traduction des gènes annotés dans EMBL, d'autres banques protéiques, de publication scientifique et de quelques soumissions d'auteurs. TrEMBL est la version protéique de la banque nucléotidique EMBL, elle contient la traduction de toutes les parties codantes annotées d'EMBL en excluant les protéines présentes dans SWISS-PROT. PIR, qui maintenant n'existe plus, fournissait des informations organisées selon des critères taxonomiques et de similarité. Enfin, HAMAP est un projet enveloppé par le groupe SWISS-PROT, son but est d'annoter automatiquement les protéines provenant des projets de séquençage des génomes microbiens. La banque contient également des collections de familles de protéines microbiennes générées par des experts et utilisées pour l'annotation automatique. Ces banques généralistes permettent donc de centraliser toutes les séquences connues. Cependant, il existe tout de même un grand nombre d'erreurs, notamment au niveau des annotations des séquences ainsi qu'une redondance des informations dans certaines banques (Wu et al., 2006).

I.2.2.2. Banques données spécialistes

Pour pallier ces inconvénients, l'augmentation exponentielle du volume, de la diversité des séquences et la diversité des études, des banques spécialisées ont été développées ; ces développements ont permis l'introduction d'informations spécifiques à chacune permettant

Recherches bibliographiques

ainsi d'avoir des banques adaptées aux besoins des utilisateurs. Elles répondent pour la plupart soit à des besoins ponctuels, soit à des besoins liés à des secteurs d'activité bien précis. Parmi celles-ci, des banques thématiques ont été développées se consacrant à un domaine bien précis. Ainsi, certaines regroupent des données sur les structures moléculaires tridimensionnelles telles que la PDB (**Berman et al., 2000**).

D'autres s'intéressent à la structure en domaine des séquences protéiques comme la banque ProDom (**Servant et al., 2002**). Il y en a également qui centralisent des données sur les signatures, caractéristiques de certaines protéines telle que PROSITE (**Hulo et al., 2006**). D'autres encore traitent des séquences et des structures d'ARN (**Gattiker et al., 2003**).

I.2.3. Différentes étapes et logiciels d'une étude phylogénétique

Il existe une myriade de logiciels, implantés sur différentes machines depuis le microordinateur jusqu'à des ordinateurs les plus puissants, écrits dans tel ou tel langage, traitant tel ou tel problème. Dans le domaine de la reconstruction phylogénétique, Dans cette section, nous allons décrire quelques grandes catégories de logiciels utilisés pour les analyses phylogénétiques (**Felsenstein, 2004**).

I.2.3.1. Logiciels de nettoyage

Tout d'abord toutes les séquences étaient expurgé, à l'aide de logiciels de nettoyage comme :

- **Sequencher**

C'est un logiciel de bio-informatique produit par la société Gene Codes Corporation, ce logiciel assemble et aligne plusieurs séquences d'ADN contiguës relativement courtes afin de créer des séquences plus longues (**Felsenstein, 2004**).

- **Bioedit**

Beaucoup de chercheurs dans le domaine de la biologie moléculaire ont utilisé des modules de BioEdit dedans pendant leur recherche originale. BioEdit a été employé pour des études moléculaires de différents organismes tels que des génomes de virus (**Ron et al., 2005 ; Chen et al., 2006**).

I.2.3.2. Logiciels d'alignement

D'une manière informelle, l'alignement de deux séquences consiste à mettre en évidence les similitudes et les différences entre les deux séquences. L'alignement de séquences a pour objectif de mettre en correspondance les portions homologues des molécules, afin de retrouver de la façon la plus cohérente possible le signal phylogénétique.

Recherches bibliographiques

Evidemment, on s'offre aussi la possibilité de sauter quelques lettres (Comet, 1998). Donc l'alignement peut être vu comme une série de transformations permettant de passer d'une séquence à l'autre (Bérard, 2003). Parmi les logiciels qui font l'alignement, on a :

- **Clustal W**

C'est un outil pour aligner la protéine multiple ou séquences de nucléotides. L'alignement est réalisé par l'intermédiaire de trois étapes : par paires alignement, génération de guide-arbre et alignement progressif. Clustal W-MPI est distribué et exécution parallèle de Clustal W. Chacune des trois étapes a été parallélisé pour réduire le temps d'exécution (Li, 2003).

- **Blast**

Basic Local Alignment Search Tool ; permet d'identifier des régions de similarité locale entre les séquences. Le programme compare des séquences nucléotidiques ou protéiques et calcule la significativité des résultats (Chen *et al.*, 2006).

- **Exonerate**

Outil d'alignement de séquences deux à deux. Il permet d'aligner à l'aide d'un alignement des séquences de nombreux modèles (Li, 2003).

I.2.3.3. Logiciels de reconstruction de l'arbre phylogénétique

- **PHYML**

PHYML est un logiciel qui implémente une nouvelle méthode de reconstruction de phylogénie à partir de séquences en utilisant le principe de maximum de vraisemblance. Cette méthode démarre avec un arbre initial fourni par l'utilisateur ou construit à partir d'un algorithme rapide basé sur les distances, puis elle améliore cet arbre à travers des réarrangements topologiques (Guindon, 2003).

- **DnaMLK**

C'est un programme d'inférence d'arbres phylogénétique, pour les séquences d'ADN, basé sur le maximum de vraisemblance avec la contrainte que les arbres inférés soient en conformité avec l'horloge moléculaire. L'horloge moléculaire stipule que toutes les feuilles de l'arbre sont équidistantes par rapport à la racine (au niveau de la longueur des branches). C'est un programme connexe à DnaML. Les mêmes suppositions au niveau du modèle sont valides (plus l'hypothèse de l'horloge moléculaire) (Felsenstein, 1993).

I.3. Etude phylogénétique

I.3.1. Définition de la phylogénie

Recherches bibliographiques

La phylogénie ou phylogenèse se définit, en général, dans l'encyclopédie comme « l'histoire de la formation et de l'évolution d'une espèce ». Le terme phylogenèse provient du grec phûlon « tribu » et genesis « origine » et il a été présenté par Haeckel dès 1860, qui l'a définie comme « l'histoire du développement paléontologique des organismes par analogie avec l'ontogénie ou histoire du développement individuel ». La phylogénie constitue un procédé pour construire des classifications d'espèces. La phylogénie moléculaire, dont certaines méthodes de reconstruction seront présentées par la suite, l'étude de l'histoire évolutive des espèces en se basant sur une portion de leur séquence moléculaire (**Haeckel, 1860**).

I.3.2. Phylogénie moléculaire

La phylogénie moléculaire utilisée pour comparer les séquences des molécules d'ADN ou de protéines des êtres vivants dans le but de déterminer les liens de parenté qui les unissent ainsi que pour appréhender leur histoire évolutive (phylogénèse). Un marqueur génétique est un caractère mesurable qui peut détecter une variation dans la séquence soit protéique, soit nucléique. Selon le niveau taxonomique étudié, il est essentiel de choisir un type de marqueur qui lui est adapté. Par exemple, si le but est de distinguer des groupes d'individus au niveau de l'espèce, les marqueurs utilisés devront être plus conservés que dans le cas des études de populations (**Beaulieu, 2007**).

Les marqueurs moléculaires concernant la molécule d'ADN elle-même, et, comme tels, sont considérés comme des mesures objectives de la variation (**Vicente et al., 2003**).

I.3.2.1. Marqueurs d'ADN nucléaire

Un certain nombre de marqueurs sont à présent disponibles pour détecter les polymorphismes d'ADN nucléaire. Dans les études sur la diversité génétique, les marqueurs les plus fréquemment utilisés sont les microsatellites. Les microsatellites sont à présent les marqueurs les plus utilisés dans les études de caractérisation génétique des animaux d'élevage (**Sunnucks, 2001**).

I.3.2.2. Marqueurs d'ADN mitochondrial

Le génome mitochondrial des animaux est formé par une petite molécule unique d'ADN circulaire dont la taille varie dans une gamme assez étroite chez les vertébrés (16- 20kb) (**Avise et al., 1987**). Sa séquence code pour 37 gènes dont 24 contiennent l'information pour une partie de la machinerie de traduction de la molécule elle-même (ARNt et ARNr), et les 13 autres pour les sous unités de la chaîne de transport des électrons. En plus, il est reconnu une

Recherches bibliographiques

région régulatrice (ou de contrôle) des processus de réplication et de transcription de l'ADNmt d'environ 0.8 Kb nommée D-loop chez les vertébrés. Par ailleurs, il est décrit la présence, dans plusieurs ADNmt, de petits fragments non codants avec des fonctions qui peuvent être aussi régulatrice (**Boore, 1999**).

I.3.3. Représentation phylogénétique

L'élaboration de technologies à haut débit serait inutile si l'on ne disposait pas de capacités d'analyse des données biologiques en croissance exponentielle. Ces données doivent se stocker dans des bases de données électroniques associées à des logiciels spécialement conçus pour la mise à jour, l'interrogation et l'extraction. Les informations doivent être facilement accessibles et flexibles aux interrogations pour faciliter l'extraction des informations pouvant être analysées pour éclaircir les voies métaboliques et le rôle des protéines et des gènes impliqués (**Rischkowsky et al., 2008**).

La bio-informatique est fondamentale pour associer les informations provenant de sources différentes et générer une connaissance nouvelle à partir de données existantes. Elle dispose également des potentialités pour simuler la structure, la fonction et la dynamique des systèmes moléculaires et est donc utile pour la formulation des hypothèses et pour la conduite du travail expérimental (**Rischkowsky et al., 2008**).

I.3.4. Arbre phylogénétique

Un arbre phylogénétique ou un dendrogramme est une représentation graphique de la phylogénie. Il exprime les liens entre les taxa sous la forme d'une succession de branchements (**Schmidt, 2003**).

Il contient quatre éléments principaux :

- **La racine** (au cas où l'arbre serait enraciné) qui indique l'ancêtre commun des espèces représentées dans l'arbre.
- **Les nœuds externes** (ou feuilles) représentent les espèces contemporaines pour lesquelles les informations ont été disponibles lors de la construction de l'arbre. Ils sont communément appelés taxons.
- **Les nœuds internes** qui représentent des ancêtres inférés, hypothétiques.
- **Les branches** (ou arêtes) de l'arbre qui montrent les relations de descendance entre les nœuds (les taxons). Ces arêtes peuvent avoir des longueurs. Ces longueurs peuvent correspondre à plusieurs informations dont, le taux de mutation, la distance génomique (**Schmidt, 2003**).

Recherches bibliographiques

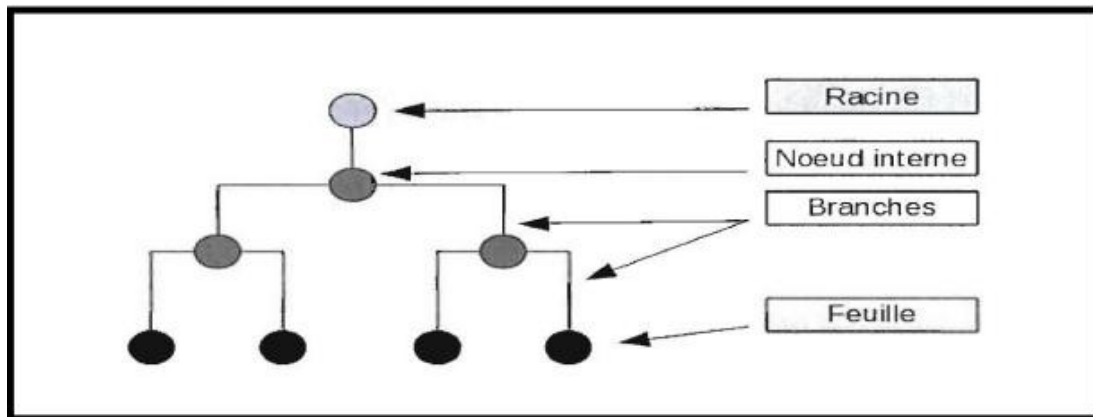


Figure 9 : Exemple d'un arbre phylogénétique (Schmidt, 2003).

I.3.4.1. Caractéristiques générales des arbres

Il est nécessaire de distinguer d'abord entre réseaux et arbres, ces derniers pouvant être non enracinés ou enracinés.

- **Un arbre enraciné** : est un arbre dans lequel un des nœuds est désigné pour être la racine, et la direction des rapports héréditaires est déterminée (Figure 10-a).
- **Un arbre non enraciné** : est une représentation intemporelle des relations phylogénétiques, les liens entre nœuds ne sont pas orientés, et un seul et unique chemin permet de passer d'un sommet à l'autre. Cet arbre n'induit donc aucune hiérarchie (Figure 10- b).

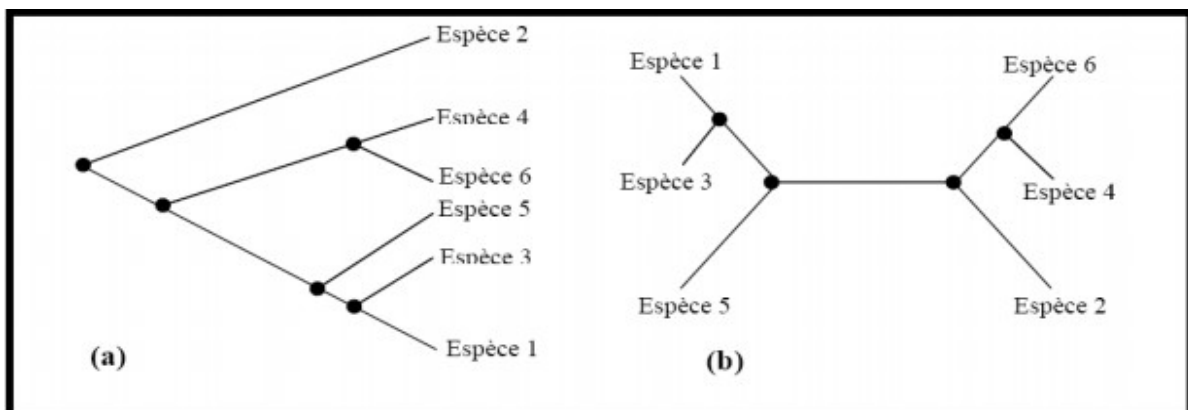


Figure 10: Les différents arbres phylogénétiques : a) arbre enraciné, b) arbre non enraciné (Schmidt, 2003).

- **Monophylie, paraphylie et polyphylie** : D'après Tourasse (1992), dans le domaine de la reconstruction phylogénétique, on distingue trois groupes taxonomiques.
- **Le groupe monophylétique** : est un ensemble d'espèces issu d'un même ancêtre commun, c'est le cas des groupe (X, Y, Z)

Recherches bibliographiques

- **Groupe paraphylétique:** c'est lorsqu'une ou plusieurs espèces d'un groupe monophylétique partagent un ancêtre commun avec des espèces appartenant à d'autres lignées, comme par exemple l'ensemble (1, 2, 3) (Figure 11)
- **Groupe polyphylétique:** si les différentes espèces d'un groupe dérivent d'ancêtres différents, celui-ci est dit polyphylétique, c'est le cas de groupe d'espèces (A, B, C Figure 11).

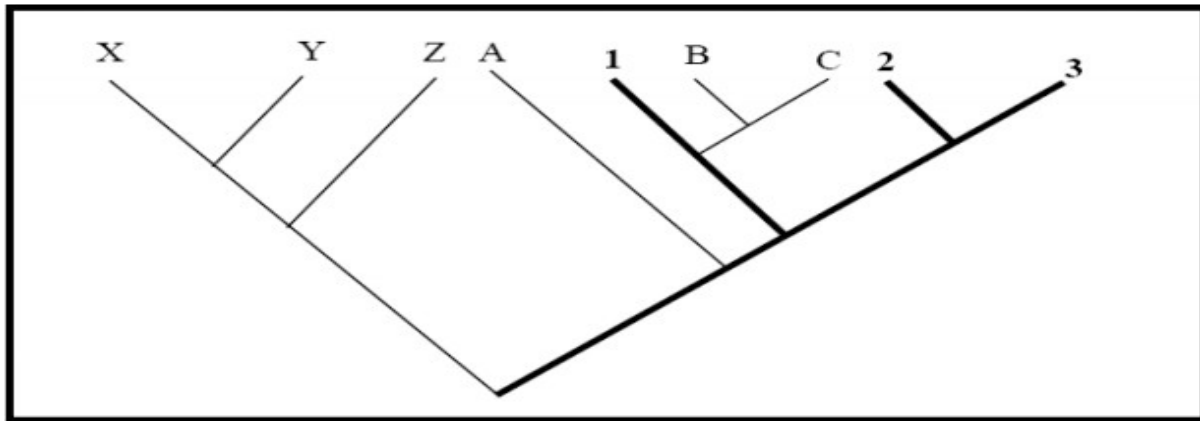


Figure 11 : Diverses catégories de groupes taxonomiques. Les taxa X, Y, Z forment un groupe monophylétique. Les taxa 1, 2 et 3 forment un groupe paraphylétique. Les taxa A, B et C forment un groupe polyphylétique (Tourasse, 1992).

I.3.4.2. Différentes représentations graphiques

Pour les arbres phylogénétiques, ils ont la capacité de véhiculer de nombreuses informations. Toutefois, avoir la capacité ne signifie pas forcément que celles-ci soient toujours exploitées. Selon la représentation graphique adoptée, il est possible de reconnaître le type d'arbre et généralement le type d'information véhiculée.

- **Le dendrogramme :** le dendrogramme est un arbre exprimant les liens entre taxons sous la forme d'une succession de branchements. Il ne désigne rien d'autre qu'un arbre dont les éléments terminaux est les taxons ou UE observés. Ce terme est assez large pour ne rien exprimer quant à la procédure utilisée pour son obtention.
- **Le cladogramme :** le cladogramme est un dendrogramme exprimant les relations phylogénétiques entre taxons et construit à partir de l'analyse cladistique où les points de branchements (les nœuds) sont définis par des synapomorphies. Ce mot a été créé la même année par (Mayr, 1965) avec des sens un peu différents.

Recherches bibliographiques

- **Les phylogrammes** : le phylogramme est un dendrogramme exprimant les branchements cladistiques et le degré de divergence adaptative subséquente aux branchements (**Mayr, 1965**).
- **Les phénogrammes** : le phénogramme est un dendrogramme produit par la taxinomie numérique où les relations entre taxons expriment les degrés de similitude globale, défini simultanément par (**Mayr, 1965**).

I.3.5.Reconstruction d'arbres phylogénétiques

La manière classique d'illustrer les relations phylogénétiques entre les espèces est de les modéliser en utilisant un arbre phylogénétique. La manière moderne consiste en la modélisation d'un réseau phylogénétique comprenant les réticulations nécessaires. A ce niveau ils présenteront seulement quelques méthodes d'inférences d'arbres phylogénétiques. La reconstruction d'un arbre phylogénétique débute par ce que l'on appelle « l'alignement » qui consiste à mettre en correspondance les sites des séquences des espèces de manière à pouvoir les comparer les unes aux autres. Les séquences utilisées pour la reconstruction peuvent être de l'ADN ou de l'ARN. Il existe trois grands types de méthodes de reconstruction d'arbres phylogénétiques: les méthodes de distances, les méthodes du maximum de parcimonie et les méthodes du maximum de vraisemblance, ces dernières sont appelées également les méthodes probabilistes (**Tourasse, 1992**).

- **L'approche phénéétique** : ne tient pas compte du processus de l'évolution. Elle se contente de mesurer les distances entre les espèces et de reconstruire le meilleur arbre possible à l'aide d'une stratégie de regroupement hiérarchique.
- **L'approche cladistique** : cherche à établir des relations de parenté en s'intéressant aux caractères (bases ou acides aminés) dérivés, partagés par les taxons. On considère ainsi tous les scénarios d'évolution en inférant les caractères des ancêtres potentiels à chaque nœud, et on choisit l'arbre qui correspond au meilleur scénario d'évolution selon un critère préalablement choisi. Les méthodes utilisées sont essentiellement basées sur le maximum de parcimonie.
- **L'approche probabiliste** : (ou maximum de vraisemblance), quant à elle, évalue en termes de probabilités l'ordre des branchements et la longueur des arêtes d'un arbre sous un - modèle évolutif donné. Les méthodes bayésiennes font aussi partie de cette approche. La première approche étudie la parenté entre les taxons en s'intéressant à leur degré de similarité alors que la deuxième est basée sur la généalogie.

MATÉRIEL
ET
MÉTHODES

Matériels et méthodes

Nous avons réalisé cette étude au niveau de l'université Saad Dahlab Blida 1 (USDB1). Ce mémoire reposant en grande partie sur l'utilisation des outils de bio-informatique dans les analyses phylogénétiques, tel que la reconstruction de l'arbre phylogénétique, dans ce chapitre, nous présenterons quelques fonctions du logiciel et bases de données durant la période de (Mars –Juillet 2022).

II.1.Extraction des données

II .1.1. Présentation de base de données NCBI

Le Centre National d'Information sur la Biotechnologie (NCBI) est la base de données des séquences de référence (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 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).

II 1.2.Etapes de la recherche sur NCBI

Toutes les séquences étudiées dans ce travail ont été téléchargées à partir de la GenBank, nous passons par plusieurs étapes distinctes :

1. Nous construirons l'URL suivante : <http://www.ncbi.nlm.nih.gov/Genbank/>.
2. Rechercher sur NCBI la séquence d'intérêt, on recherche la séquence nucléotidique souhaitée (Figure 12).
3. Nous avons choisi recherche par Nucléotide ; on sélectionne la catégorie « Nucléotide » dans le bandeau de recherche.
4. Nous avons écrit «Severe acute respiratory syndrome Coronavirus 2 » (Figure13).

Matériels et méthodes

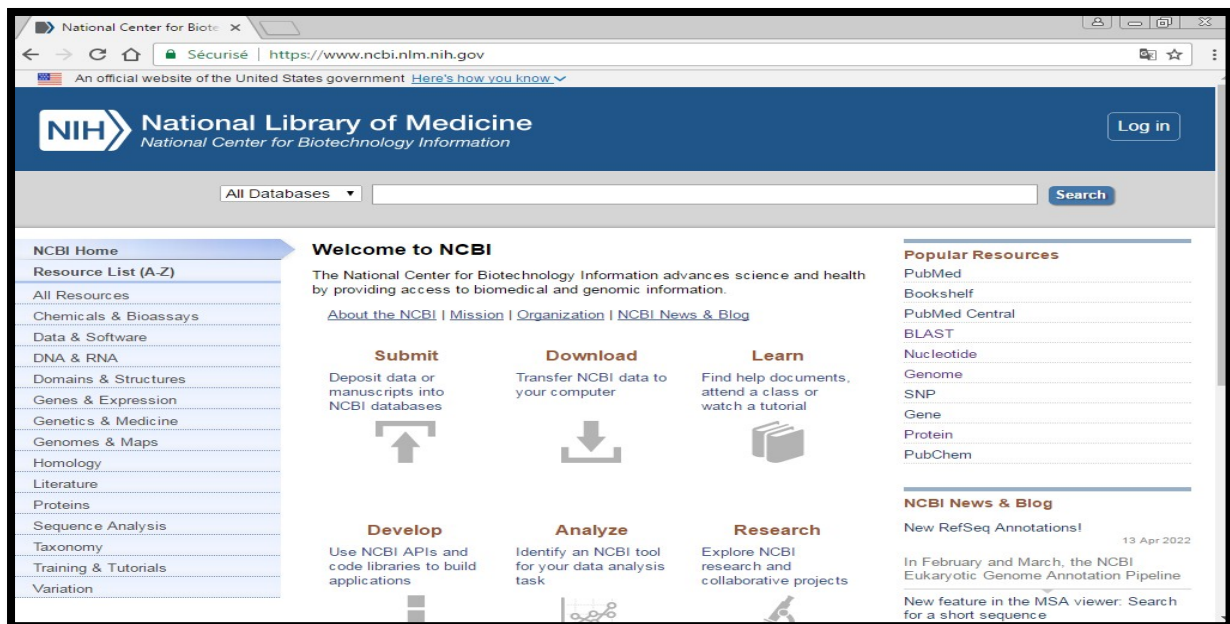


Figure12 : Page d'accueil d'NCBI (Personnelle 2022).

5. La recherche de séquences génomique s'effectue en tapant les mots clés du génome.

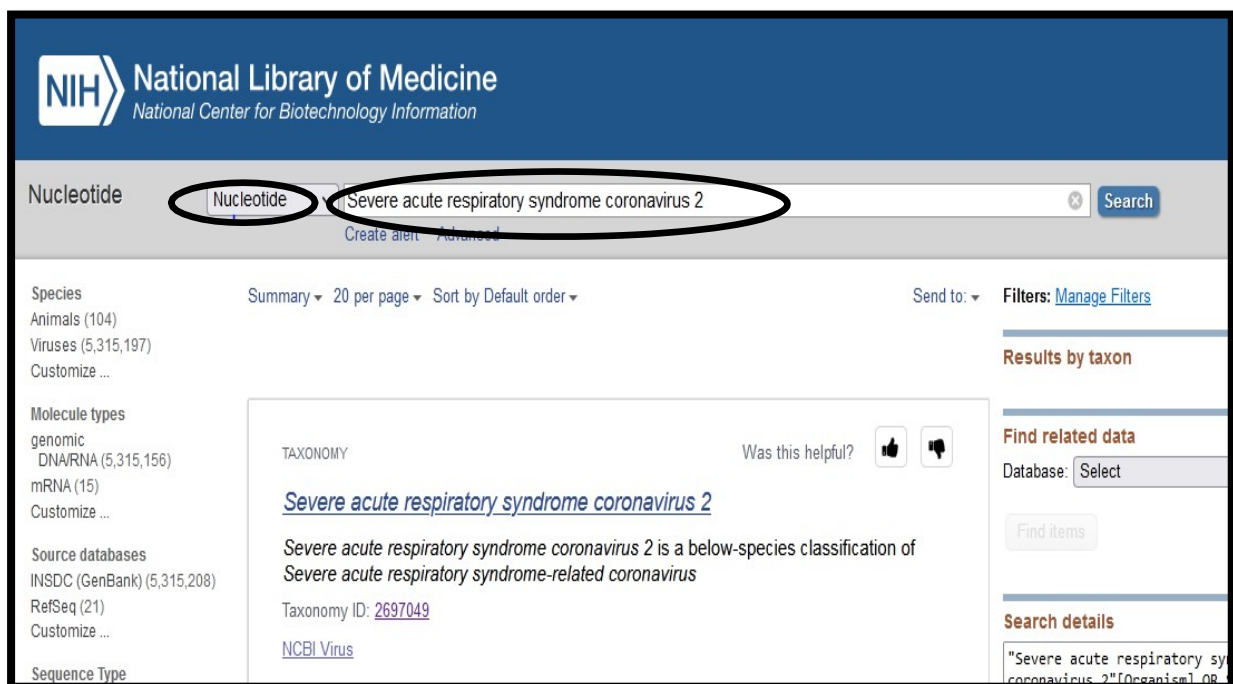


Figure 13 : Saisie des mots-clés dans la GenBank (Personnelle 2022).

6. Trier les résultats par le virus recherché (« Severe acute respiratory syndrom coronavirus 2») puis sélectionner la séquence souhaitée (Figure14).

Matériels et méthodes

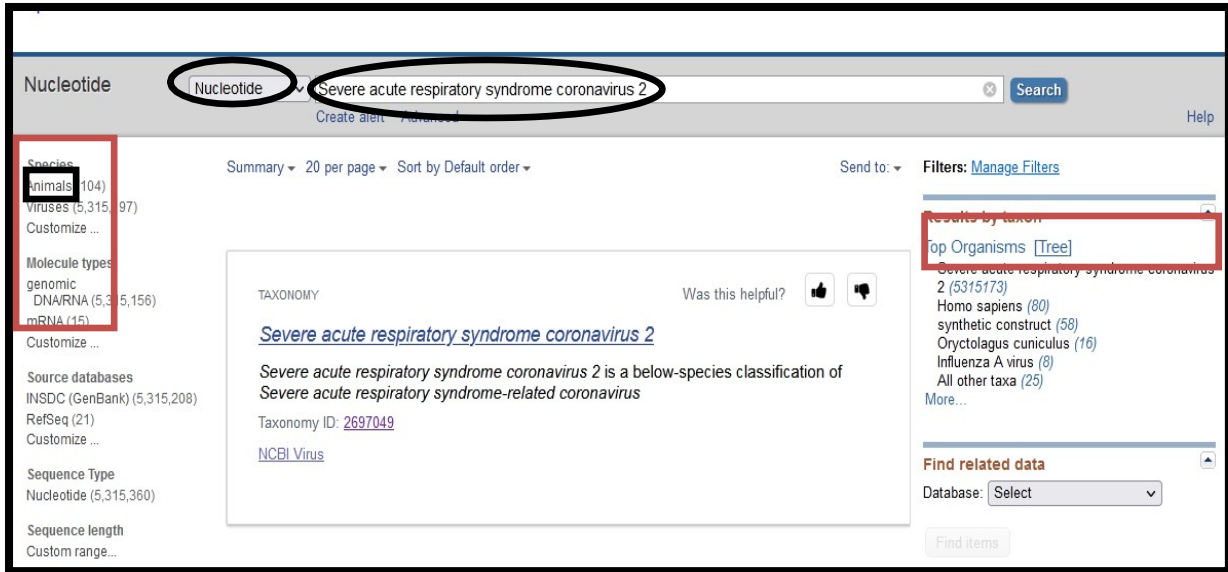


Figure 14 : Filtrage des résultats de recherche (Personnelle 2022).

7. Choisir le résultat de recherche qui correspond à notre recherche une liste de plusieurs résultats apparait, on sélectionne le résultat qui correspond à notre recherche. Dans notre cas le choix de la séquence s'est fait par rapport à la longueur de la séquence du génome (Figure 15).



Figure 15 : Sélection du résultat de recherche voulu (Personnelle 2022).

8. Une fois le lien ouvert, on voit apparaître la séquence globale du génome du SARS-COV-2 qui est divisé en plusieurs séquences qui correspondent aux domaines qui constituent le génome viral (ORF1ab, protéine S, ORF3a, N...etc.) (Figure16).

Matériels et méthodes

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1a), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E)

Identifiant séquence NCBI: 023894_NGS/2021

Type + longueur du Génome: 29787 bp RNA linear

Date de publication: 14-MAY-2022

LOCUS: ON513950.1 29787 bp RNA linear 14-MAY-2022

DEFINITION: SARS-CoV-2/human/USA/SARS00023894_NGS/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1a), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, complete cds.

ACCESSION: ON513950

VERSION: ON513950.1

KEYWORDS: .

SOURCE: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM: Severe acute respiratory syndrome coronavirus 2
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronavirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE: 1 (bases 1 to 29787)
AUTHORS: Sylvain, I., Giannikopoulos, P. and Testing Lab, I.G.I.
TITLE: Direct Submission
JOURNAL: Submitted (13-MAY-2022) Clinical Lab, Innovative Genomics Institute, 2151 Berkeley Way, Berkeley, CA 94702, USA
COMMENT: ##Assembly-Data-START##
Assembly Method: DRAGEN COVID Lineage v. v3.5.3

Figure 16 : Page du génome complet du SARS-COV-2 sur le site (Personnelle 2022).

9. Faire défiler la page jusqu'à trouver la section concernant le gène puis sélectionner l'identifiant de la séquence (Figure17).

```
/db_xref="taxon:2697049"
/country="USA"
/collection_date="2021-11-02"
gene 230..21519
/gene="ORF1ab"
CDS join(230..13432,13432..21519)
/gene="ORF1ab"
/ribosomal_slippage
/codon_start=1
/product="ORF1ab polyprotein"
/protein_id="UQN00416.1"
/translation="MESLVPGFNEKTHVQLSLPVLQVRDVLVRFVGFSDSVEEVLSEARQ
HLKDGTCGLVEVEKGVLPQLEQPYVFIKRS DARTAPHGHMVELVAELEGIQYGRSGE
TLGVLVPHVGEIPVAYRKVLLRKNKNGKAGGHSYGADLKSFDLGDDELGTDPYEDFQEN
WNTKHSSGVTRLELMRELNGGAYTRYVDNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVAVSPNECNQMLSTLMKCDHCGETSWQTG
DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPHEHSLAEYHNESG
LKTILRKGGRTIAFGGCVFSYVGCHNKCAWVPRASANIGCNHTGVVGESEGLNDNL
LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN
FKVTKGKAKKGAWNIGEQKSIILSPLYAFASEAARVRSIFSR TLETAQNSVRVLQKAA
TTTTLDGTSQYSIRLTDANNETSILATNHLVWMAVYTTGGVQLTSQHLINTEGTVYEK
```

Figure 17 : Identification NCBI de la séquence du gène (Personnelle 2022).

10. Cliquez sur le gène pour obtenir sa séquence nucléotidique. Les séquences représentatives de toutes espèces de coronavirus ont été téléchargées sous format FASTA à partir de la GenBank dans des fichiers (Figure 18). Ces fichiers peuvent être directement utilisés dans plusieurs logiciels de traitement et d'analyses des séquences.

Matériels et méthodes

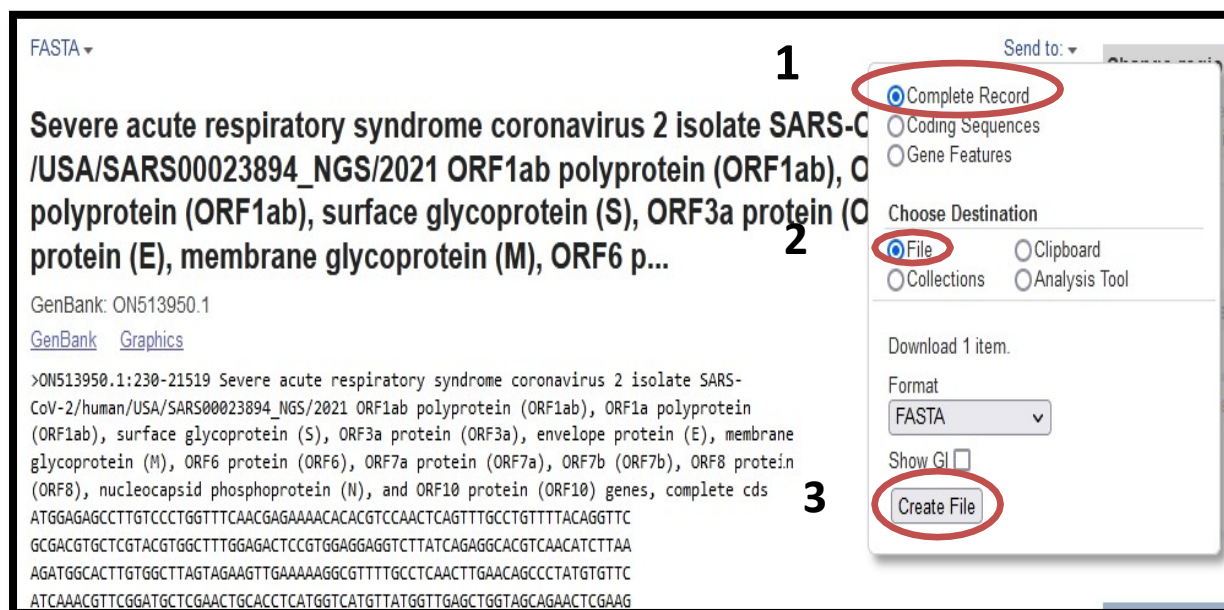


Figure 20 : Téléchargement de la séquence du gène SARS-CoV 2 sous format FASTA (Personnelle 2022).

La base de données a été construite pour sélectionner les isolats à étudier et générer les fichiers de sortie correspondant au format informatique « FASTA ». Ces fichiers peuvent alors être directement utilisés dans grand nombre logiciels de traitement et d'analyse de séquences (Dennis *et al.*, 2014).

Le format FASTA est un format textuel permettant de représenter des séquences de nucléotides ou d'acides aminés (protéines), dans lequel les nucléotides ou les acides aminés sont représentés par des codes à une lettre. Ce format permet également de faire précéder les séquences de noms de séquences et de commentaires. Cette ligne de description commence par un '>' et donne un nom ou un identifiant unique à la séquence. Elle peut également contenir des informations supplémentaires (Gabriel *et al.*, 2019).

II.2. Traitement des séquences de la GenBank

Pour réaliser une phylogénie, un bon nombre de programmes informatiques sont disponibles, pour traiter les séquences environ 92 logiciels de phylogénie et 54 web servers gratuits parmi eux : PHYLIP, PAUP*, BLAST, MEGA, ARB, FaBox (Djekoun et Hamidechi, 2010). Nous avons choisi FaBox et BLAST et le logiciel MEGA version 11, MEGA a été développé pour faciliter des analyses statistiques d'évolution moléculaire (Kamar *et al.*, 2018).

Matériels et méthodes

II 2.1. Utilisation de FaBox

FaBox (Toolbox for fastasequences) est une collection de services Web simples et intuitifs qui permet aux biologistes et aux médecins chercheurs d'effectuer rapidement une tâche typique avec des données de séquence. Les services nous facilitent la tâche pour extraire, modifier et remplacer les en-têtes des séquences et joindre ou diviser des ensembles de données en fonction de l'en-tête des informations. D'autres services incluent la fusion d'un ensemble de séquences en haplotypes et formatage automatisé des fichiers d'entrée pour un certain nombre de programmes de génétique des populations. L'utilisation de FaBox version 1.5 est une étape préliminaire et obligatoire de préparation des séquences d'ADN afin qu'elles soient traitées et analysées par les logiciels utilisés.

Nous pouvons alors procéder à l'analyse phylogénétique à partir des séquences nucléotidiques de coronavirus par différentes étapes :

1. Le programme FaBox a été employé pour convertir les fichiers, de la page principale (Figure 21). L'utilisateur peut choisir une liste de services, chacune va effectuer une tâche simple. En outre, il y a des liens aux services de conversion où l'utilisateur peut convertir n'importe quel format de données en format FASTA.

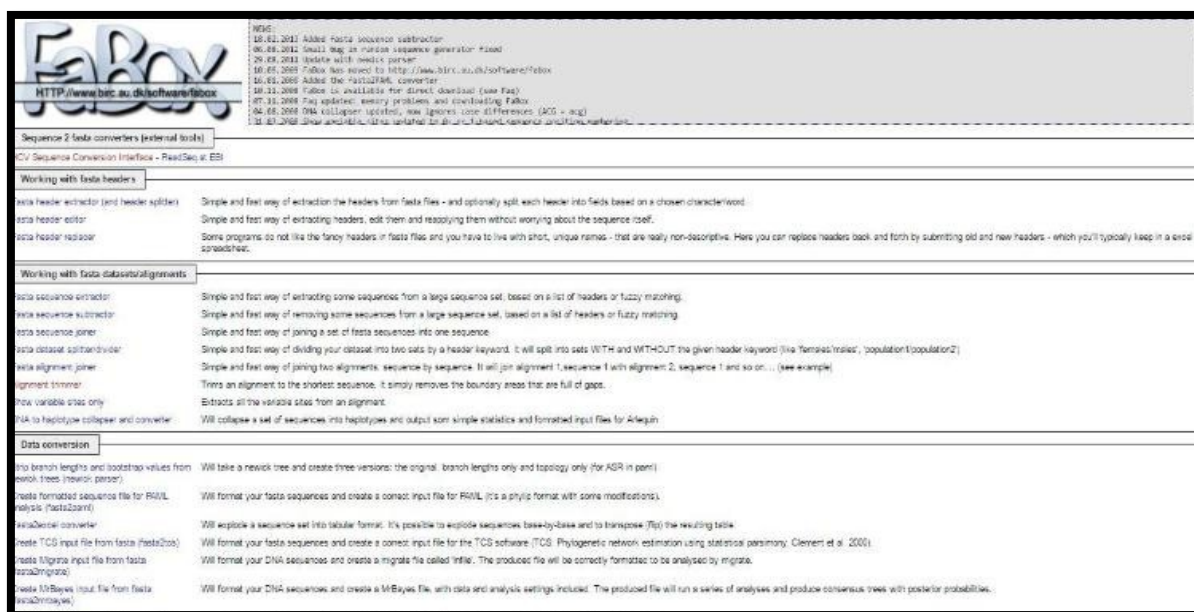


Figure 21 : La page principale du programme Fabox (Personnelle 2022).

2. L'utilisateur peut faire entrer des données en téléchargeant un fichier de données ou simplement en collant des données directement dans le web browser. On s'attend à ce que la boîte à outils se développe sur la base des demandes des services particuliers et des convertisseurs à l'avenir (Figure 22).

Matériels et méthodes

The screenshot shows the 'Format Converter' tool on the 'HCV sequence database' website. The interface includes a navigation bar with 'Housekeeping', 'Retrieving data', 'Analysis', 'Tools', 'Links', and a search box. A red box highlights the 'Input' section, which contains a text area for pasting sequences and a file upload button labeled 'Choisissez un fichier'. Another red box highlights the 'Options' section, which includes dropdown menus for 'Select input format' (set to 'Automatic') and 'Select output format' (set to 'Fasta'), along with checkboxes for 'Output line width', 'Enforce sequence name uniqueness', 'Convert Genbank to GFF3', and 'Replace IUPAC codes with Ns'. A 'Submit' button is also highlighted with a red box. Red arrows point to the 'Automatic' and 'Fasta' dropdown menus.

Figure 22 : La conversion des séquences sous forme FASTA par l'utilisation de Fabox (Personnelle 2022).

3. Alignement cropper va réduire l'alignement (en format FASTA) à la longueur de la séquence la plus courte (suppression des gaps dans le début et la fin de l'alignement). Appuyez sur "show example" pour l'essayer sans soumettre votre propre séquence (Figure 23).

The screenshot shows the 'Alignment trimmer (fasta files)' tool. It features a header with the title, a section for 'Upload or paste your fasta formatted sequences' with a 'Choose File' button and 'No file chosen' text, and an 'Action' section with 'Trim alignment' and 'Show example' buttons.

Figure 23 : Alignement trimmer (la forme FASTA) (Personnelle 2022).

Matériels et méthodes

II 2.2. Recherche des similarités par BLAST :

BLAST est une méthode de recherche heuristique utilisée en bio-informatique. Il permet de trouver les régions similaires entre deux ou plusieurs séquences de nucléotides ou d'acides aminés et de réaliser un alignement de ces régions homologues, BLAST recherche dans une base de données de séquence des segments qui sont localement homologues, BLAST utilise une matrice de similarité pour calculer des scores d'alignement. Il fournit un score pour chaque alignement trouvé et utilise ce score pour donner une évaluation statistique de la pertinence de cet alignement (Ge et al., 2020). Nous avons utilisé BLAST qui permet de retrouver rapidement dans des bases de données, les séquences répertoriées ayant des zones de similitude avec la séquence d'entrée. Cette méthode est utilisée pour trouver des relations fonctionnelles ou évolutives entre les séquences (Gabriel et al., 2019). Nous Passons par plusieurs étapes distinctes :

1. Se rendre à l'adresse suivante : <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, puis cliquer sur « Nucléotide Blast » (Figure 24).

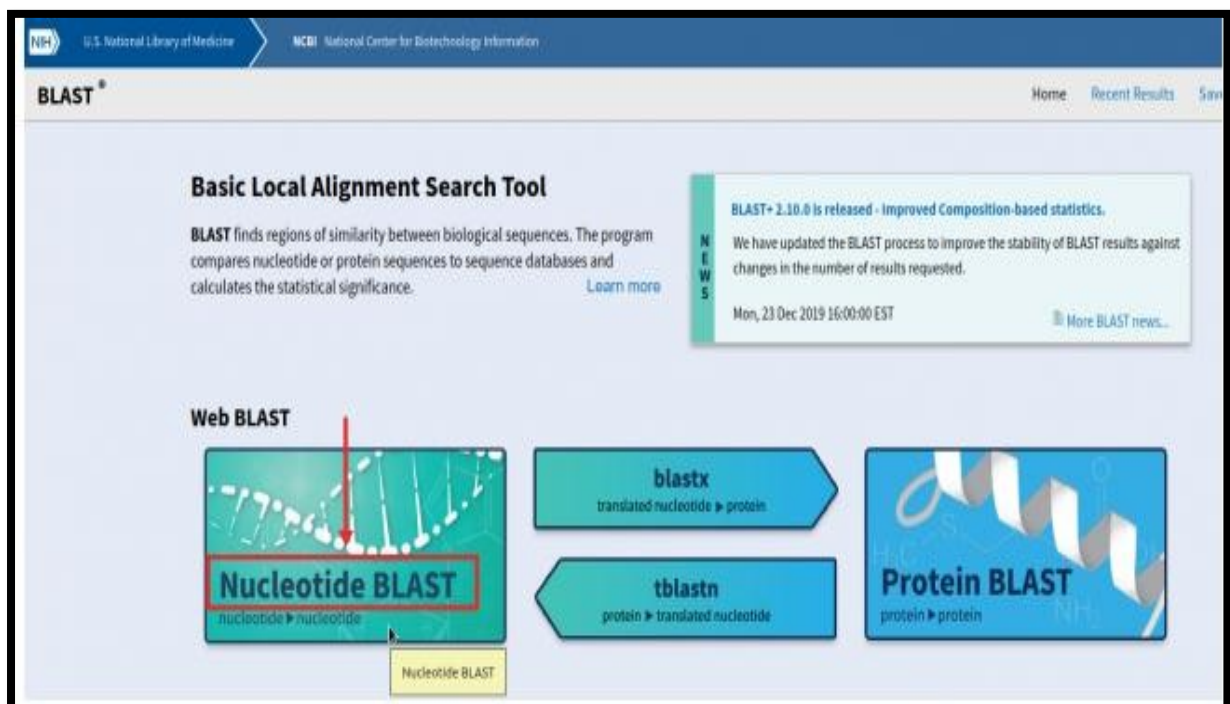


Figure 24 : Page principale de web site BLAST (Personnelle 2022).

2. Nous copions/collons la séquence recherchée dans les champs en haut de la page, nous choisissons la base de données (database), puis nous sélectionnons « Somewhat similar sequences (blastn) » et nous cliquons enfin sur le bouton « BLAST » (Figure 25,26)

Matériels et méthodes

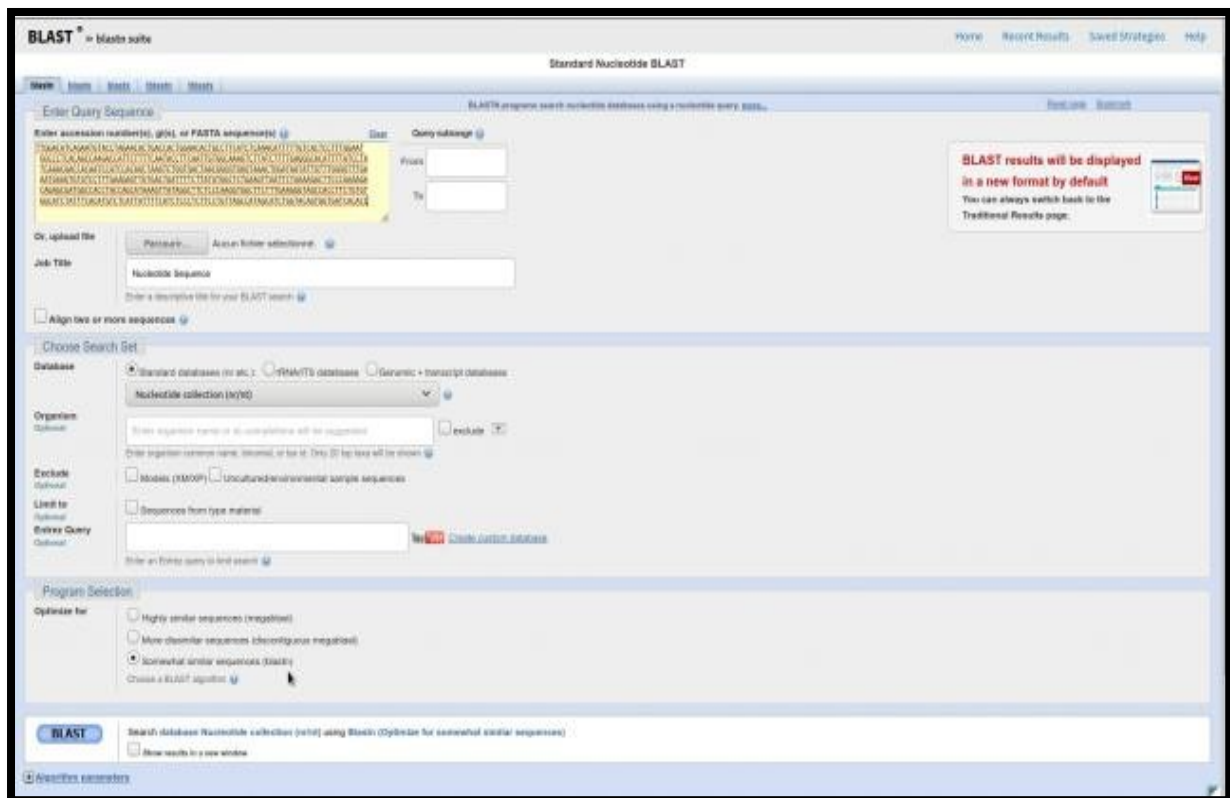


Figure 25 : Les paramètres de BLAST (Personnelle 2022).

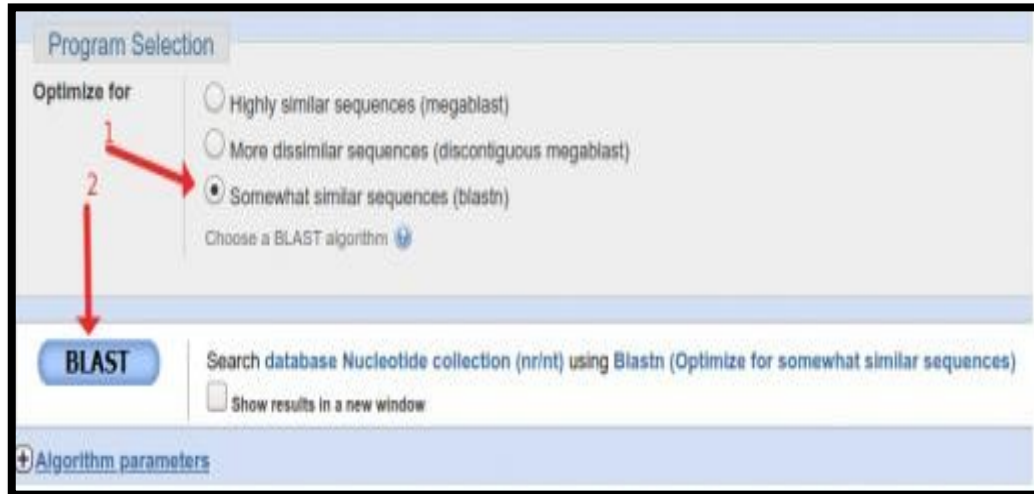


Figure 26 : Les paramètres d'algorithme de BLAST (Personnelle 2022).

3. Il faut patienter quelques secondes avant d'avoir le résultat, nous obtenons la fenêtre suivante (Figure 27) représente lecteur de la liste des résultats d'alignement.

Matériels et méthodes

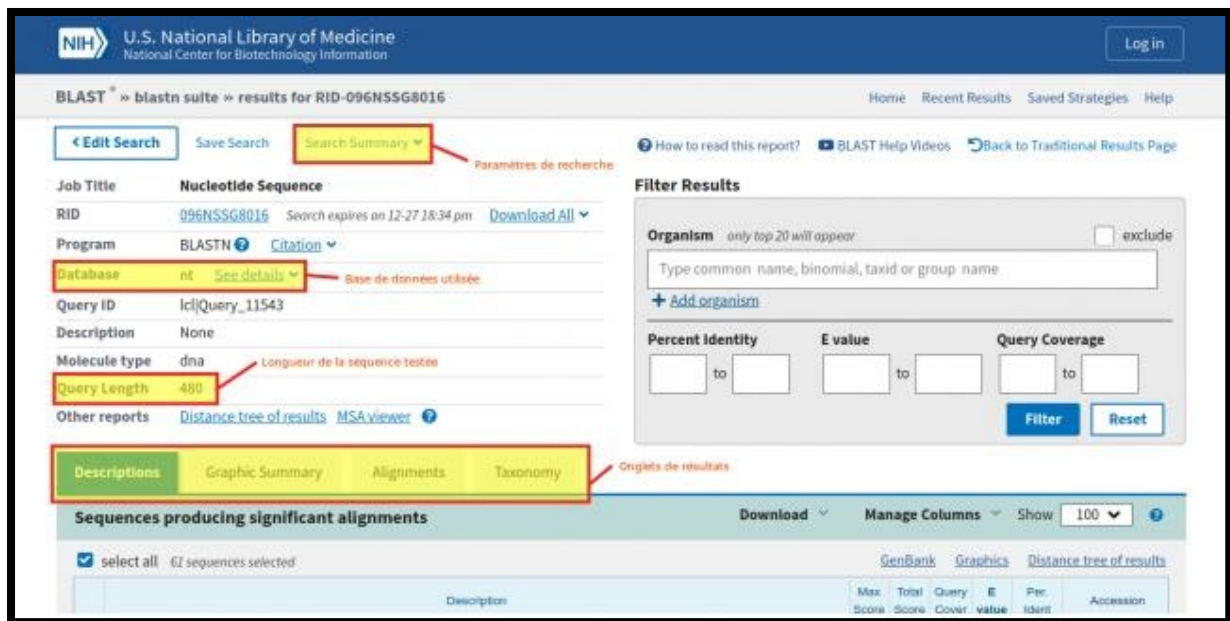


Figure 27 : Lecteur du détail des résultats d'alignement (Personnelle 2022).

4. Une fenêtre surgit avec les séquences répertoriées dans les bases de données présentant des alignements significatifs avec la séquence testée. En cliquant sur l'onglet « Graphic Summary », nous obtenons la fenêtre suivante (Figure 28).

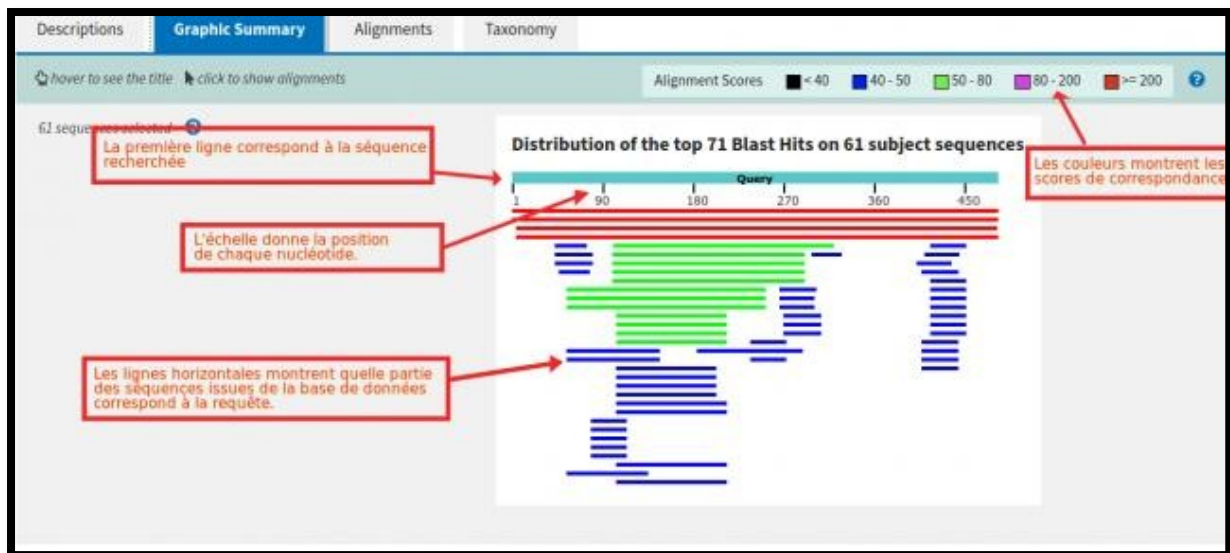


Figure 28 : Résumé graphique (Personnelle 2022).

5. En retournant sur l'onglet « Descriptions », nous allons sélectionner la séquence présentant le plus de similitudes avec notre séquence. Les résultats doivent être affichés dans un tableau

Matériels et méthodes

des séquences, ensuite nous allons choisir les séquences que nous voulons travailler avec (Figure 29).

The screenshot shows a BLAST search results page titled "Sequences producing significant alignments". The table has columns for Description, Max Score, Total Score, Query Cover, E value, Per. Ident., and Accession. Red boxes highlight the "Expected Value" column, the "Per. Ident." column, and a row with a high identity percentage. Arrows point from these boxes to descriptive text in French.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident.	Accession
Homo sapiens endothelial PAS domain protein 1 (EPAS1)	862	862	100%	0.0	97.19%	NP_018000.1
Homo sapiens BAC clone RP11-480A6 from 2...	862	862	100%	0.0	97.19%	AC016952.7
Sciurus carolinensis genome assembly, chromosome...	301	301	98%	3e-77	5.31%	LR738603.1
Sciurus vulgaris genome assembly, chromosome: 14	256	256	99%	1e-63	73.20%	LR738620.1
Mus musculus clone RP23-106L19, complete sequence	68.9	68.9	4			
PREDICTED: Vombatus ursinus tubulin monoglycosylase TTL3-like (LOC114044463), trans...	59.9	59.9	39%	2e-04	68.37%	XM_027863949.1
PREDICTED: Vombatus ursinus tubulin monoglycosylase TTL3-like (LOC114044463), trans...	59.9	59.9	39%	2e-04	68.37%	XM_027863944.1
PREDICTED: Vombatus ursinus tubulin monoglycosylase TTL3-like (LOC114044463), transcript variant X3, mRNA	59.9	59.9	39%	2e-04	68.37%	XM_027863941.1
PREDICTED: Vombatus ursinus tubulin monoglycosylase TTL3-like (LOC114044463), transcript variant X1, mRNA	59.9	59.9	39%	2e-04	68.37%	XM_027863939.1
Rhesus Macaque BAC CH250-413Q14 (r) complete sequence	53.6	53.6	4			
Homo sapiens chromosome 5, clone RP11-582J16, complete sequence	53.6	53.6	4			
Homo sapiens chromosome 5, clone RP11-382J24, complete sequence	53.6	53.6	4			
PREDICTED: Vicugna pacos leucine rich repeat and fibronectin type III domain containing 5 (LRPH5), transcript variant X5, mRNA	50.9	50.9	22%	0.10	71.82%	XM_031674999.1

Figure 29 : Etape de téléchargement des similarités forme FASTA
(Personnelle 2022).

II.2.3. Présentation de MEGA version11

Le Logiciel Molecular Evolutionary Genetics Analysis (MEGA) met en œuvre de nombreuses méthodes et outils analytiques pour la phylogénomique. Il y a une transformation de MEGA pour permettre une utilisation multiplateforme sur les systèmes d'exploitation Microsoft Windows et Linux. MEGA version11 : ne nécessite pas de logiciel de virtualisation ou d'émulation et offre une expérience utilisateur uniforme sur toutes les plateformes ; a également été mis à niveau pour utiliser plusieurs cœurs de calcul pour de nombreuses analyses évolutives moléculaires ; est disponible en deux interfaces (graphique et ligne de commande) et peut être téléchargé gratuitement sous Microsoft Windows ; destiné à être utilisé sur les systèmes 64 bits ayant la version 11 et les versions ultérieures ; multiplateforme profitera à de nombreux utilisateurs existants et permettra à de nombreux autres d'envisager d'utiliser le MEGA version11 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 recherche (Kumar et al., 2018).

Matériels et méthodes

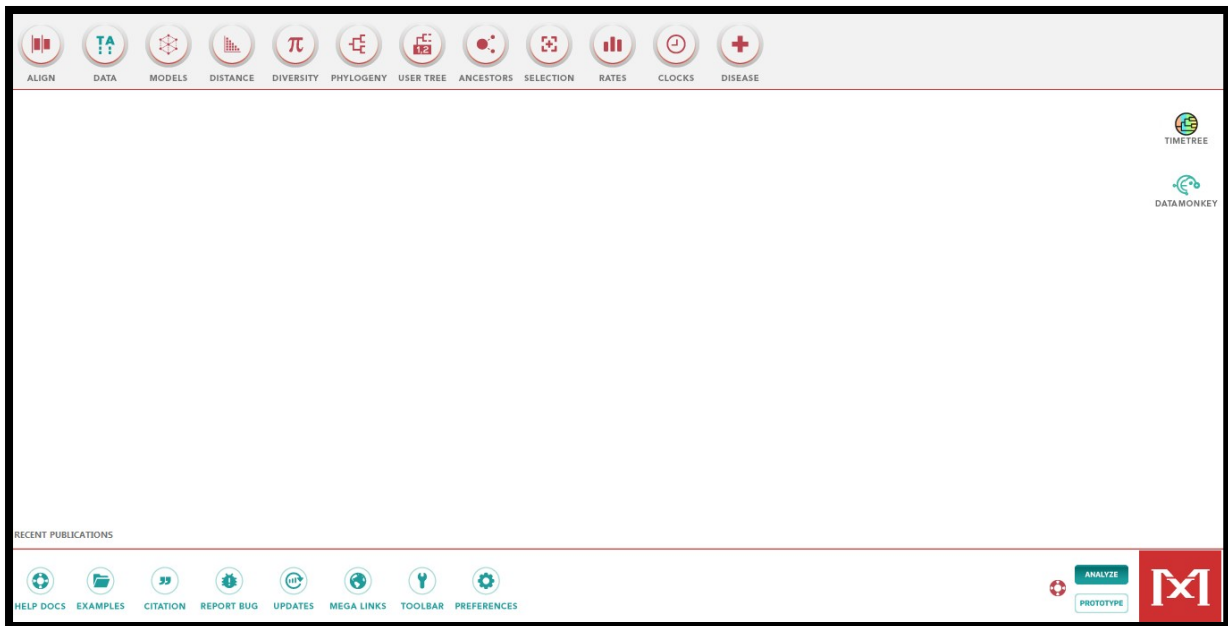


Figure 30 : La fenêtre principale de MEGA11 (Personnelle 2022).

II.2.3.1. Alignement des séquences

Plusieurs types d'alignement consistent à mettre en évidence les similitudes et les différences entre les séquences, nous avons choisi l'alignement multiple car il permet également de mettre en évidence les zones conservées dans un ensemble de séquences. Nous Passons par plusieurs étapes distinctes :

1. Lancez l'explorateur d'alignement en sélectionnant Aligner / Edit /Build Alignment dans la barre de lancement de la fenêtre principale de MEGA (Figure 31).

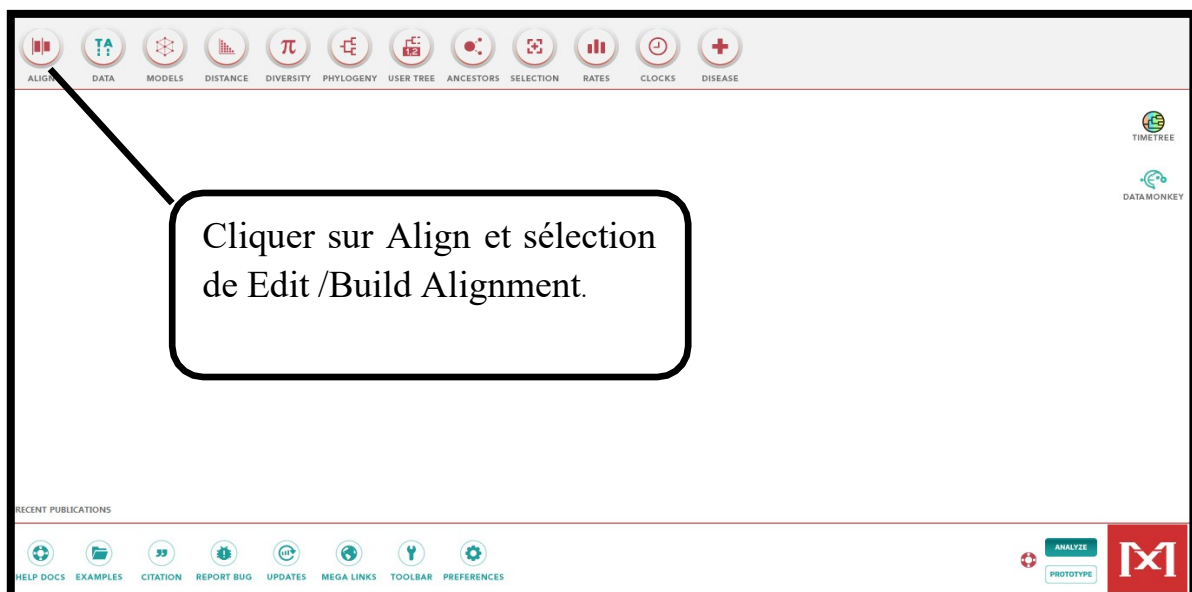


Figure 31 : Ouvrir le programme de MEGA11. Cliquer sur Align et sélection de Edit /Build Alignment (Personnelle 2022).

Matériels et méthodes

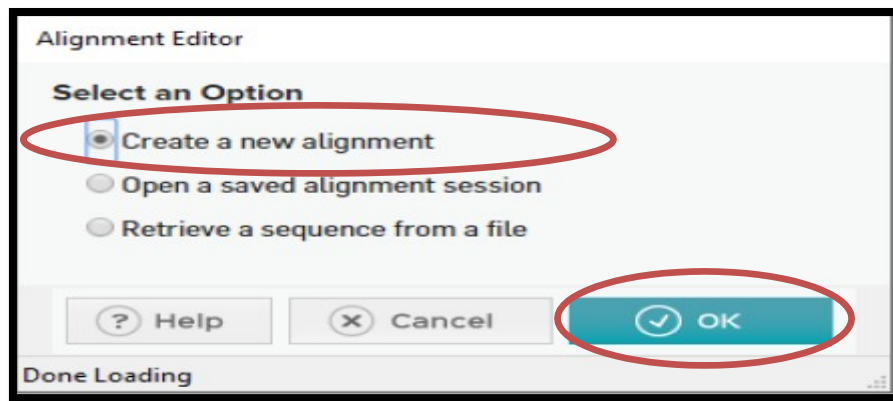


Figure 32: Choisir create a new alignment et cliquer sur OK (Personnelle 2022).

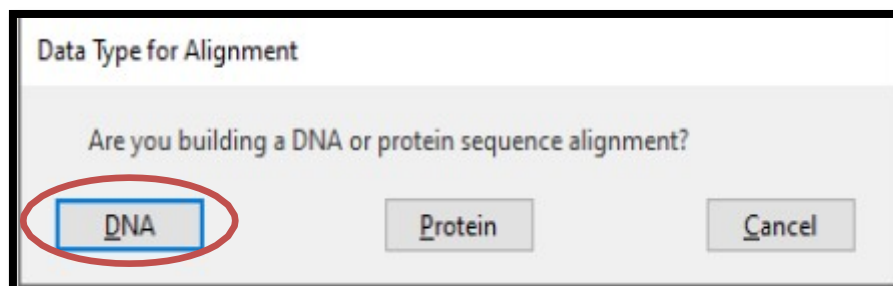


Figure 33: Cliquer sur DNA dans MEGA11 (Personnelle 2022).

2. La fenêtre qui apparaît ensuite permet l'introduction des séquences d'intérêt en cliquant sur Editer / Insérer une séquence vierge et en insérant Cov / le nom du pays /et le variant (Figure 34).

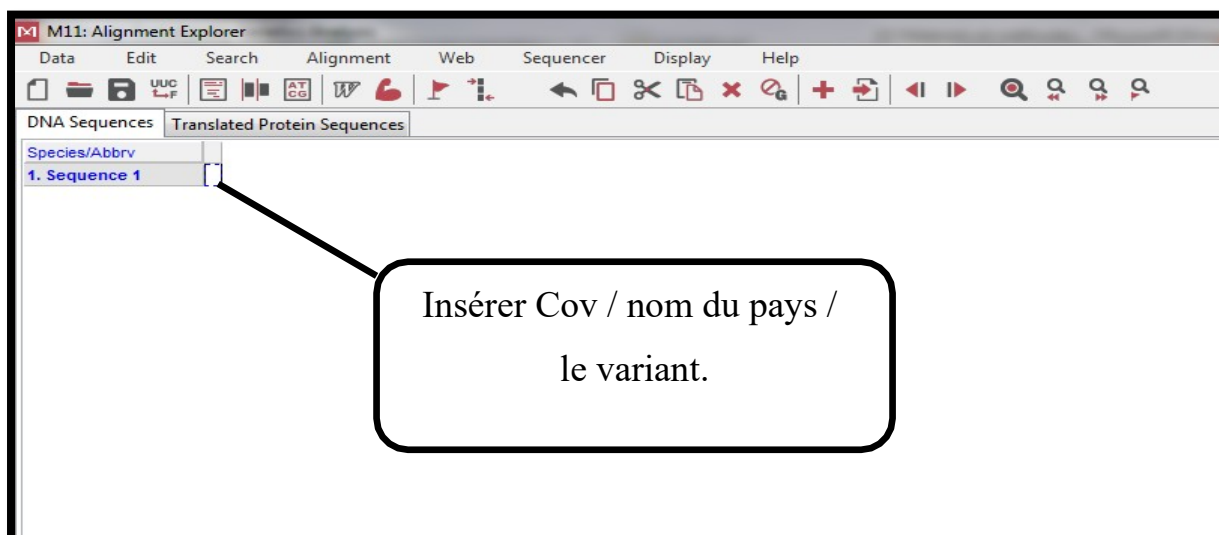


Figure 34 : Cliquer sur séquence et écrire Cov, nom de pays, le variant dans MEGA11 (Personnelle 2022).

3. Cliquer sur Edit et choisir insert sequence from File (Figure 35).

Matériels et méthodes

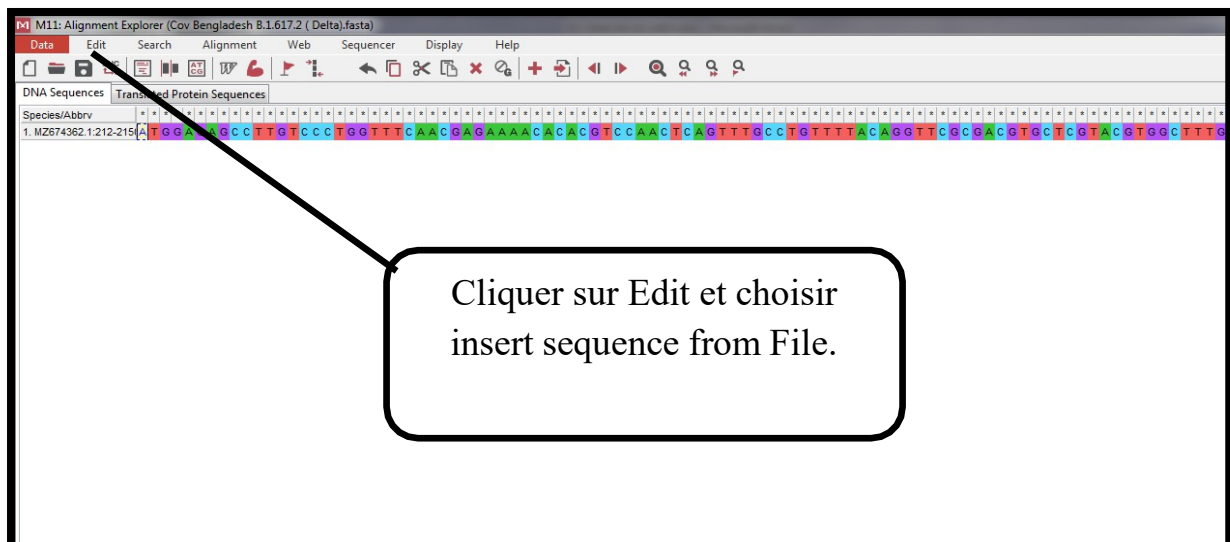


Figure 35 : Insertion de séquences ADN de SARS-CoV-2 dans MEGA (**Personnelle 2022**).

4. Insertion des échantillons d'ADN du Virus SARS-CoV-2 obtenus d'NCBI dans MEGA11 (Figure 36).

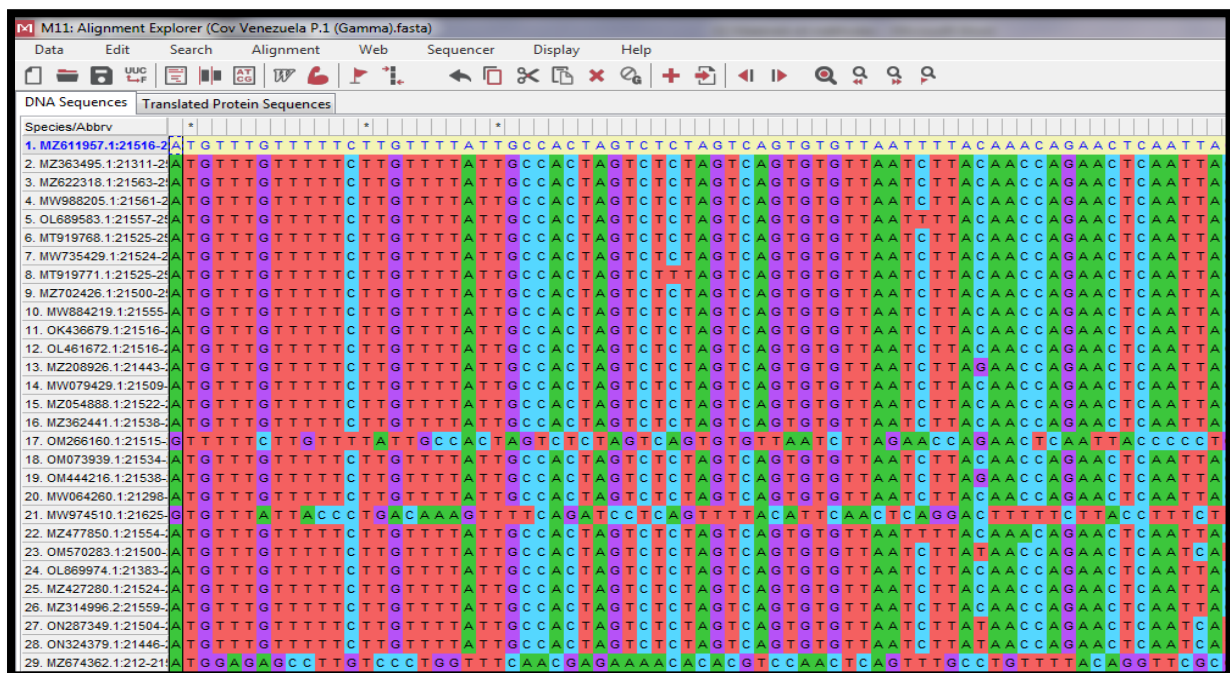


Figure 36 : Résultat d'insertion des échantillons d'ADN du Virus SARS-CoV-2obtenus d'NCBI dans MEGA11 (**Personnelle 2022**).

5. Choix d'alignement, un alignement multiple a été réalisé en utilisant le programme ClustalW. Sélectionnez Alignement et choisir Align par ClustalW (Figure 37). Nous avons choisi ce type d'alignement parce qu'il est très utile et permet d'étudier les différences et les similitudes entre les séquences d'un même marqueur moléculaire (**Ivoet *al.*, 2004**).

Matériels et méthodes

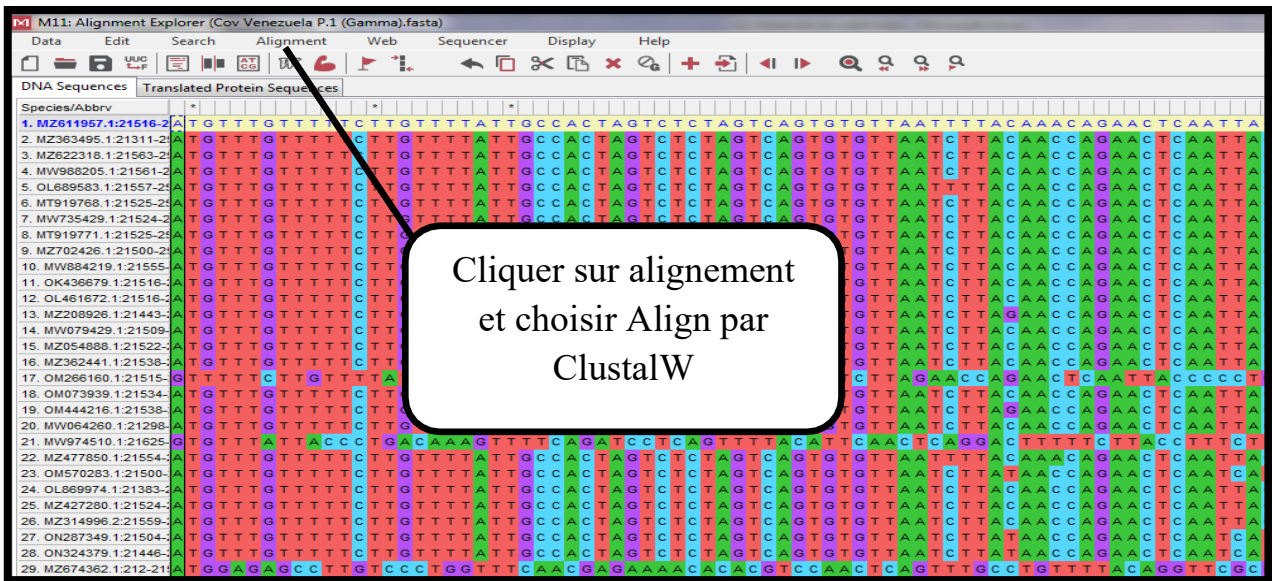


Figure 37: Sélectionner Edit / select all pour sélectionner toutes les séquences sur MEGA11 (Personnelle 2022).

6. Enregistrement des résultats, 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 (Figure 38).

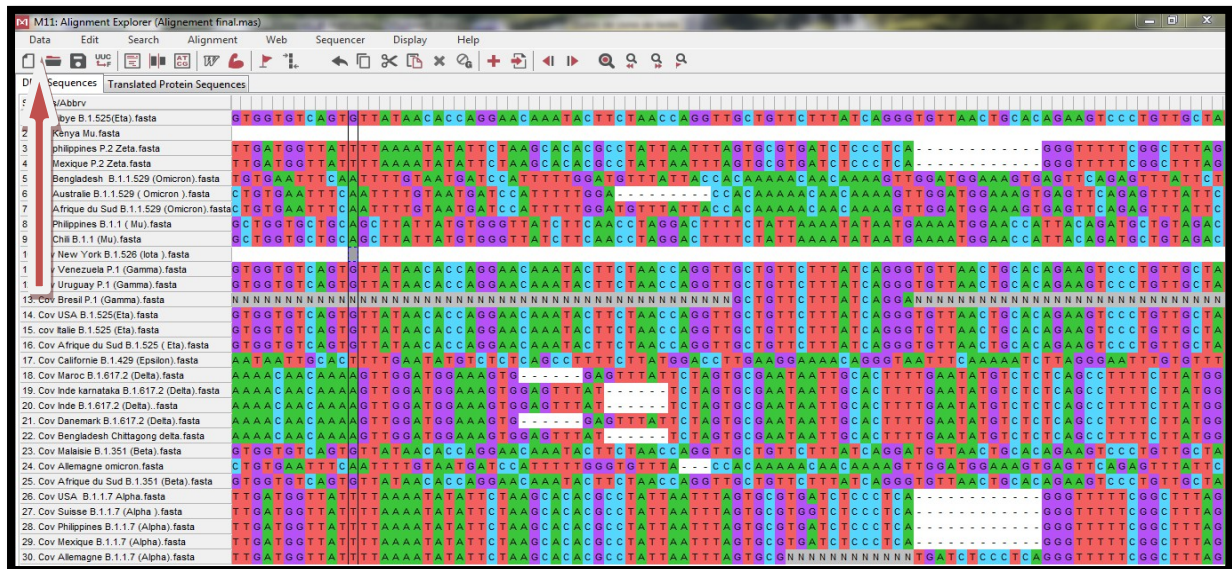


Figure 38 : Enregistrement des résultats de l'alignement (Personnelle 2022).

II.2.3.2. Construction des arbres phylogénétiques

Plusieurs paramètres de la construction d'arbre phylogénétique sont montrés dans cette étape, nous avons choisi les meilleurs paramètres : "UPGMA", "Bootstrap", "Kimura 2 paramters".

Matériels et méthodes

1. Dans la barre de lancement principale de MEGA11, cliquer sur phylogeny puis Contract / Test UPGMA Tree (Figure 39). Nous avons choisi la méthode UPGMA car c'est une méthode utilisée pour la construction de l'arbre phylogénétique, plus simple, basée sur l'hypothèse que les taux de mutation et donc les vitesses d'évolution sont identiques sur les différentes branches de l'arbre (Kumar *et al.*, 2018).



Figure 39 : Barre de lancement principale de MEGA11 (Personnelle 2022).

2. Test de robustesse des phylogrammes (Bootstrap), dans cette étape les transitions et les transversions ne sont pas équivalentes en termes de proportions et leur fiabilité a été évaluée par le programme "Bootstrap" qui a été appliqué avec un nombre de répétitions égal à 1000. Le test Bootstrap mesure la cohérence interne d'un ensemble de données moléculaires en analysant si des alignements légèrement modifiés supportent les mêmes clades. Plus précisément, il s'agit d'un rééchantillonnage test afin qu'à chaque cycle un alignement répété soit construit (Lutteropp *et al.*, 2019).

3. Choix de modèle, en utilisant le modèle à deux paramètres de Kimura(1980) corrige les occurrences multiples, en tenant compte des taux de substitution transitionnelle et transversionnelle, tout en supposant que les quatre fréquences nucléotidiques sont les mêmes et que les taux de substitution ne varient pas entre les sites (Lutteropp *et al.*, 2019).

4. Substitutions à inclure, MEGA contient plusieurs méthodes statistiques pour estimer l'évolution entre les séquences en fonction du nombre de différences observées. Ces méthodes comprenaient la méthode correcte pour les substitutions multiples en tenant compte des biais de transition/transversion, des fréquences de base inégales, des taux de substitution variables entre les sites (Lutteropp *et al.*, 2019).

Matériels et méthodes

-Cliquer sur OK pour lancer la construction (Figure 40).

M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope →	All Selected Taxa
Statistical Method →	UPGMA
PHYLOGENY TEST	
Test of Phylogeny →	Bootstrap method
No. of Bootstrap Replications →	1000
SUBSTITUTION MODEL	
Substitutions Type →	Nucleotide
Genetic Code Table →	Not Applicable
Model/Method →	Kimura 2-parameter model
Fixed Transition/Transversion Ratio →	Not Applicable
Substitutions to Include →	d: Transitions + Transversions
RATES AND PATTERNS	
Rates among Sites →	Uniform Rates
Gamma Parameter →	Not Applicable
Pattern among Lineages →	Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment →	Pairwise deletion
Site Coverage Cutoff (%) →	Not Applicable
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
SYSTEM RESOURCE USAGE	
Number of Threads →	1

? Help X Cancel ✓ OK

Figure 40 : Les paramètres utilisés pour la reconstruction phylogénétique (Personnelle 2022).

RÉSULTATS
ET
DISCUSSION

Résultats et Discussion

Dans ce chapitre nous comparons les différents variants du SARS-COV-2 par les résultats obtenus dans la base de données GenBank et traités par le logiciel MEGA11, ainsi que nous déterminons la localisation des mutations du coronavirus dans les arbres phylogénétiques de quatre gènes étudiés qui sont S, ORF1ab, N et ORF3a.

III.1. Résultats

III.1.1.GenBank

Le traitement de 5353266 séquences d'ADN de SARS-COV-2 publiées sur la GenBank nous a permis de classer les 4 gènes les plus variable de coronavirus, les gènes sont étudiés dans 30 pays différents du monde, qui sont soit des gènes structuraux N (Nucléocapside protéine), S (Spike glycoprotéine), ou bien des gènes fonctionnels ORF (Open Reading Frame):ORF1ab (ORF1a et ORF1b), ORF3a (**Derouiche et al., 2021**),les résultats obtenus sont représentés dans le tableau 3.

Tableau 3 : Séquences nucléotidique de différents variants de Coronavirus extraites de laGenBank.

Résultats et Discussion

Gènes	Mutations	pays	Variants	
S	A570D, T716I	Kenitra, Maroc	B.1.1.7 (Alpha)	
	P681R	Dhaka, Bangladesh		
	D1118H , L730F, T1001	Lima, Pérou		
	Délé H 69 /V70	Philippines		
	S982A	Manille, philippines		
	E484K, D614G	Mexique		
	R408I	Delhi, Inde		
	D178H, L452R	USA		
	N501Y, déléh 69 /V70	Berlin, Allemagne		
	P681H	New York		
	D614G	Allemagne		
	E484K, D614G	Suisse		
	P681H, K417T	Genève, Suisse		
	K417N, E484K, N501Y	Afrique du sud		B.1.351 (Beta)
	E484K	Allemagne		
	D614G	Malaisie		
	P681R	Mohakhali, Bangladesh	B.1.617.2 (Delta)	
	E156G, déléh F157	Chittagong, Bangladesh		
	T19R, G142D, E156GT478K, D614G D950N	Inde		
	L452R	Delhi, Inde		
	E484K	Danemark		
	T19R, G142D, E156G	Kolkata, Bangladesh		
	T19R, G142D, R158G, L452R, T478K, K417T, D614G, P681R, D950N	Maroc		
	Délé H 69 /V70	Kerala, Inde		
	D950N,T95I	Egypte		
	D614G	Venezuela		P.1 (Gamma)
	E484K	Uruguay		
	N501Y	Allemagne		
	E661D	Brésil		
	G339D, N440K S477N, T478K	Afrique du sud	B.1.1.529 (Omicron)	
	DéléH 69 /V70, N501Y	Bengladesh		
	Q493R, G496S, Q498R, N501Y	Australie		
S371P ,S373P ,S477A	Rabat, Maroc			
G142D, N211I, G339D,S371L, S373P, S375F, G446S	Chine			

Résultats et Discussion

	N764K, D496Y N856K, H655Y	Belgique	
ORF1ab	P472L	Malaisie	B.1.1.7 (Alpha)
	P4804P , L5784A	Kenitra, Maroc	
	T1001I	Dhaka, Bangladesh	
	T1001I, A4489V, P4619 L	Hyderabad, Inde	
	I2230T, T1001I, P314L	New York	
	N764K	Liban	
	S1433P,P472L, D6909G	Pérou	
	P4619 L, P4715L	Bengladesh	
	T1001I, I2230T P314L	Mexique	
	T58I , L37F, P314L	Israël	
	A4489V	Bengladesh 2	
	L2780F, R1383K	Thaïlande	
	I2230T, M429I	Inde	
	M902I	USA	
	A1708D, P2908I M902I ,T1001I	Delhi, Inde	
	L2780F	Lima, Pérou	
	T1001I, I2230T, A1708D	Colombie	B.1.351 (Beta)
	P309L,P314L, P4619 L	Chine	
	T1001I, P2046L P314L,I2230T,P4619 L	Egypt	B.1.617.2 (Delta)
	I2230T	Casablanca, Maroc	
S370L ,K977 Q L71F,P323L	Uruguay	P.1 (Gamma)	
T1001I	Venezuela		
P314L, R226K	Brésil		
T3255I, T1001I P314L, P4619L	Belgique	B.1.1.529 (Omicron)	
Y655T, L856A P924P, I2230T	Rabat, Maroc		
N	S235F	Kenitra, Maroc	B.1.1.7 (Alpha)
	S235P	Lima, Pérou	
	P314L	Liban	
	S235F ,R203K G204R	New York	
	G204R	Mexique	
	Y73C	Dhaka, Bangladesh	
	G204R	Chittagong, Bangladesh	
	L139F	Thaïlande	
	P344S	Hyderabad, Inde	

Résultats et Discussion

	P13L D3L	USA	B.1.617.2 (Delta)	
	L37F,N1410T, A116V	Royaume-Uni		
	P344S	Pérou		
	D63G, P344S	Egypte		
	A211V, T141I G204R	Mohakhali, Bengladesh		
	R203k	Bengladesh		
	D377Y	Maroc		
	P80R	Singapour		
	G215C	Mexique		
	D63G	Thaïlande		
	R203K K204R	Venezuela		P.1 (Gamma)
	P80R, P344S	Brésil		
	R203K, G204R	Morelos, Mexique		
	D63G, R203K G215C, D377Y	Australie		B.1.1.529 (Omicron)
	R203K,G204R	Rebat, Maroc		
	P240S	Belgique		
ORF3a	W128L, L140V G251V	Delhi, Inde	B.1.1.7 (Alpha)	
	S26L	Maharashtra, Inde		
	S171L, G251V, S26L	Philippines		
	L275F	Mexique		
	I82T	New York		
	G251V	Royaume-Uni		
	Q57H , W131C W193L , T223I	Israël		
	Q57H	Inde		
	F15L	Thaïlande		
	Q57H	Inde		B.1.351 (Beta)
	S171L, Q57H, S26L, G251V	USA		
	Q57H	USA	B.1.617.2 (Delta)	
	Q57H, S26L	Egypte		
	Q57H, G251V	Danemark		
	S26L	Mexique		
	S26L	Maroc		
	S253P, Q57H	Uruguay	P.1 (Gamma)	
	S253P, S26L, G251V	Brésil		
	T64T, Q57H, G251V	Rabat, Maroc	B.1.1.529 (Omicron)	
	S92L, D155Y, S26L	France		

Résultats et Discussion

	T1001I, A1708D, I2230T	Colombie	B.1.351 (Beta)	
	P309L,P314L, P4619 L	Chine		
	T1001I, P2046L P314L,I2230T,P4619 L	Egypt	B.1.617.2 (Delta)	
	I2230T	Casablanca, Maroc		
	S370L ,K977 Q L71F,P323L	Uruguay	P.1 (Gamma)	
	T1001I	Venezuela		
	P314L, R226K	Brésil		
	T3255I, T1001I P314L, P4619L	Belgique	B.1.1.529 (Omicron)	
	Y655T, L856A P924P, I2230T	Rabat, Maroc		
N	S235F	Kenitra, Maroc	B.1.1.7 (Alpha)	
	S235P	Lima, Pérou		
	P314L	Liban		
	S235F ,R203K G204R	New York		
	G204R	Mexique		
	Y73C	Dhaka, Bangladesh		
	L139F	Thaïlande		
	P344S	Hyderabad, Inde		
	P13L D3L	USA		
	L37F,N1410T, A116V	Royaume-Uni		
	P344S	Pérou	B.1.617.2 (Delta)	
	D63G, P344S	Egypte		
	A211V, T141I G204R	Mohakhali, Bangladesh		
	R203k	Bangladesh		
	D377Y	Maroc		
	P80R	Singapour		
	G215C	Mexique		
	D63G	Thaïlande		
	R203K K204R	Venezuela		
	P80R, P344S	Brésil		
	R203K, G204R	Morelos, Mexique	P.1 (Gamma)	
	D63G, R203K G215C, D377Y	Australie		
	R203K,G204R	Rebat, Maroc	B.1.1.529 (Omicron)	
	P240S	Belgique		
		W128L, L140V G251V	Delhi, Inde	
		S26L	Maharashtra, Inde	

Résultats et Discussion

ORF3a	S171L, G251V, S26L	Philippines	B.1.1.7 (Alpha)
	L275F	Mexique	
	I82T	New York	
	G251V	Royaume-Uni	
	Q57H, W131C W193L, T223I	Israël	
	Q57H	Inde	B.1.351 (Beta)
	F15L	Thaïlande	
	Q57H	Inde	B.1.617.2 (Delta)
	S171L, Q57H, S26L, G251V	USA	
	Q57H	USA	P.1 (Gamma)
	Q57H, S26L	Egypte	
	Q57H, G251V	Danemark	
	S26L	Mexique	
	S26L	Maroc	B.1.1.529 (Omicron)
	S253P, Q57H	Uruguay	
	S253P, S26L, G251V	Brésil	
	T64T, Q57H, G251V	Rabat, Maroc	
	S92L, D155Y, S26L	France	

III.1.2. Analyse phylogénétique

Le logiciel MEGA 11 nous a permis de construire 4 arbres phylogénétiques pour 4 gènes différents (N, S, ORF1ab, ORF3a), chaque gène est représenté par plusieurs séquences nucléotidiques selon le pays. Le tableau 4 nous donne un aperçu de la structure des différents arbres en définissant le nombre exact de branches, nœuds et feuilles pour chaque arbre phylogénétique obtenu.

Tableau 4 : Résultats des arbres phylogénétiques obtenus par MEGA des gènes de SARS Cov-2.

Racine	Branches	Nœuds	Feuilles
S	26	4	35
ORF1ab	20	7	25
N	20	3	25
ORF3a	21	2	20

III.1.3. Construction des arbres phylogénétiques par MEGA11

Afin d'établir les relations phylogénétiques entre les différents variants de Coronavirus nous avons étudié les quatre gènes les plus variables (Figure 41) :

Gène S : La protéine S est une protéine transmembranaire qui facilite la liaison de l'enveloppe virale aux récepteurs de l'enzyme de conversion de l'angiotensine 2 « ACE2 » situés à la surface des cellules hôtes. Structuellement la protéine S est une très grande protéine transmembranaire qui s'assemble en trimères pour former des pointes de surface (SPIKE) distinctives des coronavirus, elle est divisée en deux parties fonctionnellement distinctes ; sous unité S1 et S2, la sous unité S1 contient un domaine de liaison au récepteur RBD « Receptor Binding Domain » qui est responsable de la reconnaissance et de la liaison avec le récepteur de surface cellulaire(ACE2). La sous-unité S2 est la "tige" de la structure qui assure la fusion des membranes virales et cellulaires lors de réarrangements structurels importants **(Fung et al., 2020)**.

Gène N : Les nucléocapsides N des Bêta-coronavirus partagent une structure de domaine globale commune. C'est une phosphoprotéine de 43-50 kDa qui s'associe à l'ARN génomique pour former la nucléocapside hélicoïdale, elle participe aussi à la réplication de l'ARN, à la formation des virions et à l'évasion immunitaire **(Fung et al., 2020)**. 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)**.

Gène ORF1ab : contient un domaine RdRp spécifique jouant un rôle central dans la transcription et la réplication de l'ARN virale **(Naqvi et al., 2020)**. Il exprime une polyprotéine ORF1ab 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 structurales ORF1a et ORF1ab partagent la majorité du contenu nucléotidique du génome virale. Les polyprotéines pp1a et pp1ab sont codées par les ORF 1a et ORF1b **(Naqvi et al., 2020)**.

Gène ORF3a : est un gène trouvé dans les coronavirus du sous-genre Sarbecovirus, y compris le SARS-CoV et le SARS-CoV-2. Il code pour une protéine accessoire d'environ 275 résidus d'acides aminés de long, dont on pense qu'elle fonctionne comme une viroporine, c'est la plus grande protéine accessoire et a été la première des protéines accessoires du SARS-CoV à être décrite. C'est une protéine transmembranaire qui contient trois domaines transmembranaires.

Résultats et Discussion

Il possède un ectodomaine N-terminal et un endodomaine C-terminal, qui est séparé du domaine transmembranaire par une région riche en cystéine, ce domaine il fonctionne comme un dimère ou un tétramère, qui est assemblée au niveau de la membrane plasmique. Il peut également former des oligomères d'ordre supérieur, avec des effets fonctionnels inconnus (Kern *et al.*, 2021).

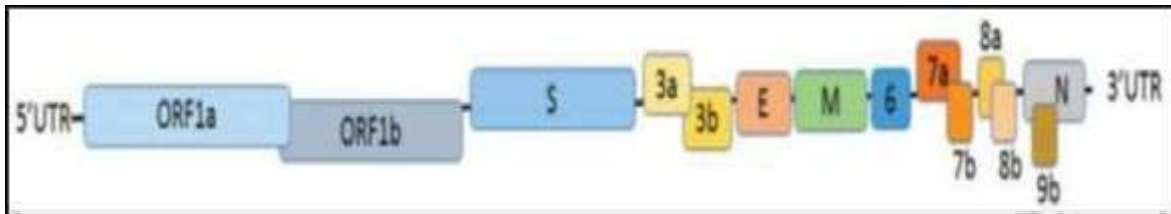


Figure 41 : Organisation génomique du SARS-CoV-2 (Segondy 2020).

III.1.3.1. Arbre phylogénétique du gène de pointe S

L'arbre phylogénétique (Figure 42) a été réalisé sur la base du gène S de différents variants du Coronavirus qui ont touché plusieurs pays du monde avec une taille qui varie entre 4374pb et 1992pb. Avec l'émergence rapide de plusieurs variants, l'OMS a établi un système de classification pour distinguer les variants émergents du SARS-CoV-2, plusieurs variants du SARS-COV-2 ayant en commun les mêmes mutations sont apparus dans plusieurs pays tels que ; le variant Anglais (Alpha B.1.1.7), les variants sud-africain (Béta B.1.351) et (Omicron B1.1.529), le variant Brésilien (Gamma P.1) et le variant indien(Delta B.1.617.2), des études ont montré que les variants tendent à se propager plus vite car ils sont plus transmissibles ou plus infectieux, mais ne semblent pas entraîner une maladie plus grave ou une mortalité plus élevée ou des symptômes différents de ceux observés depuis le début de la pandémie (OMS, 2021).

Résultats et Discussion

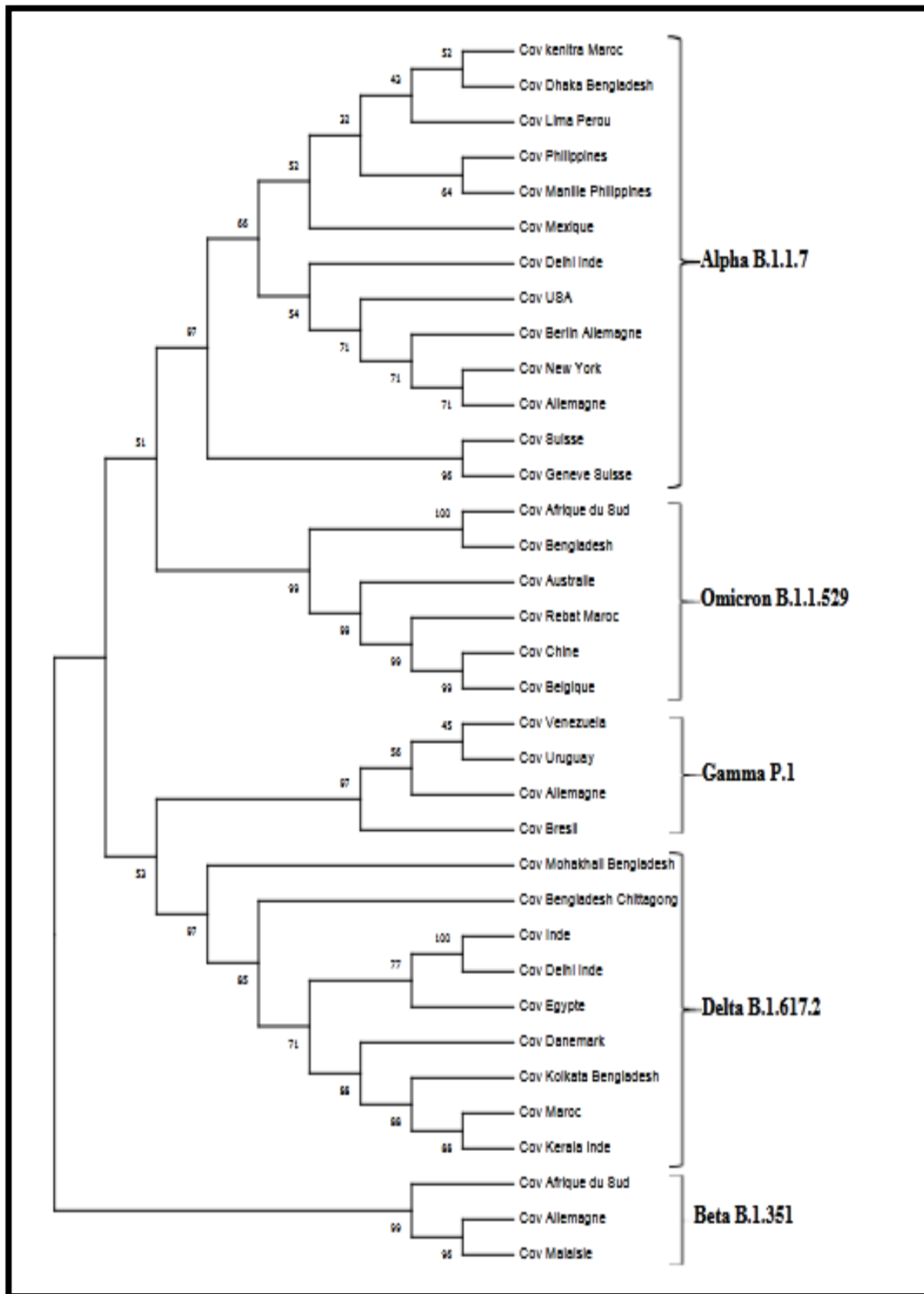


Figure 42 : Arbre phylogénétique de gène S représentant différents variants de Coronavirus.

Résultats et Discussion

- Les différents groupes présentés dans les clades de l'arbre du gène S sont :

-Alpha (B.1.1.7)

Le groupe du variant Alpha contient 13 séquences qui ont presque la même taille et qui contiennent différentes mutations dans plusieurs pays dont : La séquence de Kenitra (Maroc) avec une taille de 21529 jusqu'à 25341 pb (A570D, T716I), Dhaka (Bengladesh) de 21500 – 25325 pb (P681R), Lima (Pérou) de 21384-25196 pb (D1118H, L730F, T1001), Philippines de 21320-25486 (Délé H 69 /V70), Manille (Philippines) de 21524-25336 pb (S982A), Mexique de 21516-25328pb (E484K, D614G), Delhi (Inde) de 21550-25368 pb (R408I), USA de 21563-25384(D178H, L452R), Berlin (Allemagne) de 21510-25330pb (N501Y, Délé H69 /V70), New York de 21320-25360 pb (P681H), Allemagne de 21559-25392 pb (D614G), la Suisse de 21557-25378 pb (E484K, D614G), Genève (Suisse) de 21503-25315 pb (P681H, K417T).

-Beta (B.1.351)

Le groupe du variant Beta contient 3 séquences avec des tailles et des mutations différentes dans les pays suivants : La taille de la séquence Afrique du Sud varie entre 21504-25316 pb (K417N, E484K, N501Y), Allemagne de 21524- 25336 pb (E484K), Malaisie de 21509 - 25330 (D614G).

-Delta (B.1.617.2)

Les séquences du variant Delta représentent les mutations et les tailles selon les pays suivants : La séquence Mohakhali (Bengladesh) varie entre 21561-25376pb (P681R), Chittagong(Bengladesh) de 21509-25324 pb (E156G, Délé F157), Inde de 21515 25324 pb (T19R, G142D, E156G, T478K, D614G, D950N), Delhi (Inde) de 21223-25250 (L452R), Danemark de 21538-25353pb (E484K), Kolkata (Bengladesh) 21525-25355pb (T19R, G142D, E156G), Maroc de 21443-25258pb (T19R, G142D, R158G, L452R, T478K, D614G, P681R, D950N), Kerala (Inde) de 21529-25344 pb (Délé H 69 /V70), Egypte de 21534-25355 pb (D950N, T95I).

- Gamma (P.1)

Les séquences du variant Gamma comprennent les tailles et les mutations dans les pays suivants: La taille de la séquence de Venezuela varie entre 21516-25337 pb (D614G), Uruguay de 21561-25382 pb (E484K), Allemagne de 21552-25364 (N501Y), Brésil de 21543-25322 pb (E661D).

Résultats et Discussion

- Omicron (B.1.1.529)

Le groupe du variant Omicron regroupe des séquences de différents pays avec des tailles et des mutations distinctes qui sont : La séquence sud-africaine varie entre 21446-25258 pb (G339D, N440K, S477N, T478K), Bangladesh de 21500-25312 pb (DéléH 69 /V70, N501Y), Australie de 21383-25816 pb (Q493R, G496S, Q498R, N501Y), Maroc de 21526-25338 pb (S371p, S371P, S373P, S477A), Chine de 21497-25309 pb (G142D, N211I, G339D, S371L, S373P, S375F, G446S), Belgique de 21497-25309 pb (N764K, D496Y, N856K, H655Y).

La protéine S compte 30 mutations dans le tableau 5, réparties sur la totalité de son code génétique, parmi ces mutations celles situées dans le domaine RBD « K417N/T, L452R, K478T, E484K, E484Q et N501Y » sont celles qui ont eu le plus d'impact sur la virulence du virus, sur sa transmissibilité ainsi que sur sa capacité à échapper au système immunitaire, par conséquent les variants qui abritent ces mutations sont les plus dominants et les plus contagieux au monde (Nelson et al., 2021).

Tableau 5 : Caractéristiques des différentes mutations qui ont touchées les séquences présentées dans l'arbre du gène S.

Résultats et Discussion

Gène	Mutations	Base changée et sa position dans le génome	Type de mutation	Impact de la mutation sur la protéine
S	G142D	Guanine est remplacée par adenine en position 425	Transition	Remplacement de Glycine(G) en acide aspartique (D) à la position 142
	T19R	Cytosine est remplacée par guanine en position 56	Transversion	Remplacement de thréonine(T) en arginine (R) à la position 19
	D950N	Guanine est remplacée par adenine en position 2848	Transition	Remplacement de thréonine(D) en arginine (R) à la position 950
	L452R	Thymine est remplacée par guanine en position 1355	Transversion	Remplacement de thréonine(T) en arginine (R) à la position 452
	D614G	Adenine est remplacée par guanine en position 23403	Transition	Remplacement de l'acide aspartique (D) en glycine (G) à la position 614
	P681R	Cytosine est remplacée par guanine en position 23604	Transversion	Remplacement de proline (P) en arginine (R) à la position 681
	E484K	Guanine est remplacée par adenine en position 1450	Transition	Remplacement de l'acide glutamique(E) en lysine(K) à la position 484
	E156G	Adenine est remplacée par guanine en position 467	Transition	Remplacement de l'acide glutamique (E) en glycine(G) à la position 156
	T95I	Cytosine est remplacée par thymine en position 284	Transition	Remplacement de thréonine (T) en isoleucine(I) à la position 95
	N501Y	Adenine est remplacée par thymine en position A23063T	Transversion	Remplacement d'asparagine (N) en tyrosine (Y) à la position 501
	D178H	Guanine est remplacée par cytosine en position 532	Transversion	Remplacement de l'acide aspartique(D) en histidine(H) à la position 178
A570D	Cytosine est	Transversion	Remplacement d'alanine	

Résultats et Discussion

	remplacée par adenine en position 23271		(A) en acide aspartique(D) à la position 570
T716I	Cytosine est remplacée par thymine en position 2147	Transition	Remplacement de thréonine (T) en isoleucine (I) à la position 716
L730F	Cytosine est remplacée par thymine en position 2188	Transition	Remplacement de leucine (L) en phénylalanine (F) à la position 730
T1001I	Cytosine est remplacée par thymine en position 3003	Transition	Remplacement de thréonine (T) en isoleucine(I) à la position 1001
F888L	Cytosine est remplacée par guanine en position 2664	Transversion	Remplacement de phénylalanine (F) en leucine à la position 888
Q52R	Adenine est remplacée par guanine en position 155	Transition	Remplacement de glutamine(Q) en arginine (R) à la position 52
A67V	Cytosine est remplacée par thymine en position 200	Transition	Remplacement d'alanine (A) en valine (V) à la position 67
R408I	Guanine est remplacée par thymine en position 1223	Transversion	Remplacement d'arginine (R) en isoleucine (I) à la position 408
G339D	Guanine est remplacée par adenine en position 1016	Transition	Remplacement de glycine(G) en acide aspartique(D) à la position 339
N211I	Adenine est remplacée par thymine en position 632	Transversion	Remplacement d'asparagine(N) en isoleucine (I) à la position 211
S371L	Cytosine est remplacée par thymine en position 1112	Transition	Remplacement de sérine (S) en leucine(L) à la position 371
G339D	Guanine est remplacée par adenine en position 1016	Transition	Remplacement de glycine (G) en acide aspartique (D) à la position 339
S373P	Thymine est remplacée par	Transition	Remplacement de sérine (S) en proline(P) à la

Résultats et Discussion

	cytosine en position 1117		position 373
G446S	Guanine est remplacée par adenine en position 1336	Transition	Remplacement de glycine (G) en sérine(S) à la position 446
H655Y	Cytosine est remplacée par thymine en position 1963	Transition	Remplacement d'histidine (H) en tyrosine(Y) à la position 655
S477A	Thymine est remplacée par guanine en position 1429T	Transversion	Remplacement de sérine (S) en alanine (A) à la position 477

III.1.3.2. Arbre phylogénétique de gène ORF1ab

L'arbre phylogénétique (Figure 43) a été réalisé sur la base des séquences du gène ORF1ab de différents variants de Coronavirus qui ont touché plusieurs pays du monde avec une taille qui varie entre 4405pb et 7096pb. Comme pour l'arbre du gène S, l'arbre du gène ORF1ab est constitué de plusieurs variants, le variant Anglais (Alpha B.1.1.7), les variants sud-africain (Béta B.1.351) et (Omicron B.1.1.529), le variant indien (Delta B.1.617.2) et le variant brésilien (Gamma P.1) qui représentent des mutations apparues dans plusieurs pays.

Résultats et Discussion

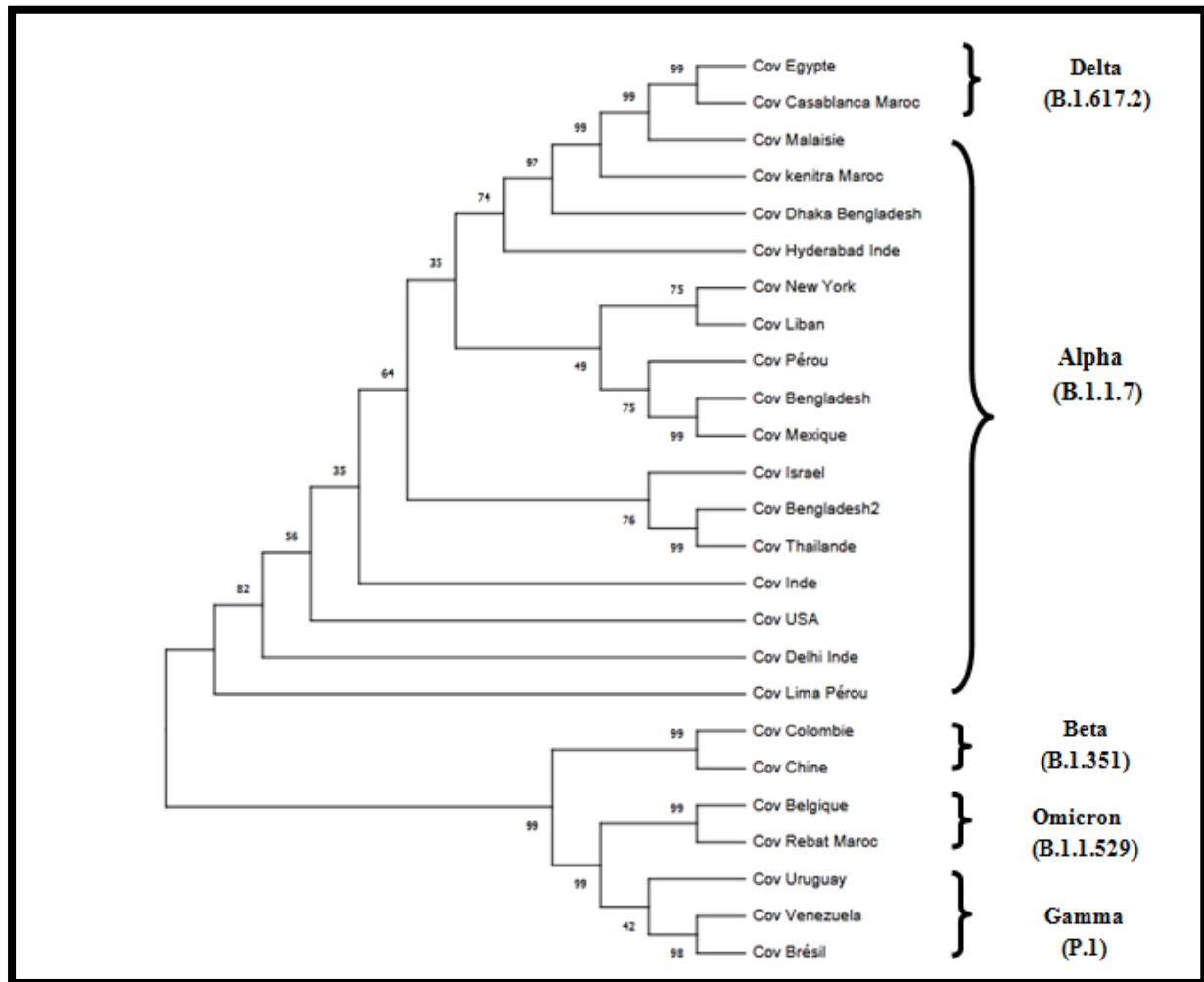


Figure 43 : Arbre phylogénétique de gène ORF1ab représentant les séquences de différents variants de Coronavirus.

Les séquences de différents variants présentés dans l'arbre du gène ORF1ab sont :

- **Alpha (B.1.1.7) :** la taille de la séquence de Malaisie varie entre 225-13530pb, Kenitra au Maroc (241-13449 pb), Dhaka Bangladesh (256 - 13464 pb), Hyderabad Inde (264-13481pb), New York (215 - 13423 pb), Liban (220- 13256 pb), Pérou (212 - 13501 pb), Bangladesh (215-13423 pb), Mexique (228 - 21508 pb), Israël (264-13481pb), Bangladesh 2 (250-13265), Thaïlande (234 - 13523 pb), Inde (244 - 13421 pb), USA (230-13423), Delhi Inde (253 - 21542 pb), Lima Pérou (225-13420). Ces séquences contiennent les mutations suivantes: Malaisie (P472L), Kenitra Maroc (P4804P, L5784A), Dhaka Bangladesh (T1001I), Hyderabad Inde (T1001I, A4489V, P4619 L), New York (I2230T, T1001I, P314L), Liban (N764K) , Pérou (S1433P, P472L, D6909G), Bangladesh (P4619 L, P4715L) , Mexique (T1001I, I2230T, P314L), Israël (T58I, L37F, P314L), Bangladesh 2 (A4489V),

Résultats et Discussion

Thaïlande (L2780F, R1383K), Inde (I2230T, M429I), USA (M902I), Delhi Inde (A2708, P2908I, M902I, T1001I), Lima Pérou (L2780F).

- **Beta (B.1.351)** : la taille de la séquence colombienne varie entre 251 - 21540 pb, la Chine (212 -13417 pb). Ces séquences ont comme mutation : Colombie (T1001I, A1708D, I2230T) et la Chine (P309L, P314L, P4619 L).

- **Delta (B.1.617.2)** : la taille de la séquence égyptienne varie entre 234 -13454 pb, Casablanca Maroc de 241 - 21530 pb, celles qui ont les mutations : Egypte (T1001I, P2046L, P314L, I2230T, P4619 L), Casablanca Maroc (I2230T).

- **Gamma (P.1)** : la taille de la séquence Uruguay varie entre 264- 21553 pb, Venezuela de 228 - 21508 pb, Brésil de 272-21546 pb, Ces séquences ont les mutations suivantes : Uruguay (S370L, K977Q, L71F, P323L), Venezuela (T1001I), Brésil (P314L, R226K).

- **Omicron (B.1.1.529)** : la taille de la séquence de Belgique varie entre 212 - 13417 pb, Rabat Maroc de 241 - 13446pb. Ces séquences contiennent les mutations suivantes : Rabat, Maroc (Y655T, L856A, P924P, I2230T), Belgique (T3255I, T1001I, P314L, P4619 L).

La protéine ORF1ab compte 30 mutations dans le tableau 6, réparties sur la totalité de son code génétique, parmi ces mutations celles situées dans le domaine ORF1ab (I2230T, P314L, T1001I, P4619L). D'après **Selene et al. (2022)** ces mutations sont les plus graves, elles affectent la virulence du virus (transmissibilité, évasion, résistance aux vaccins et médicaments).

Tableau 6 : Caractéristiques des différentes mutations des séquences que nous avons présenté dans l'arbre du gène ORF1ab.

Gène	Mutations	Base changée et sa position dans le génome	Type de mutation	Impact de la mutation sur la protéine
Orflab	L2780F	Thymine est remplacée par cytosine en position 832T	Transition	Remplacement de leucine (L) en phénylalanine (F) à la position 2780
	R1383K	Adénine est remplacée par thymine en position 4012	Transversion	Remplacement De arginine (R) en lysine (K) à la position 1383
	P314L	Cytosine est remplacée par	Transition	Remplacement de proline (P) en leucine

Résultats et Discussion

	thymine en position 15324		(L) à la position 314
S1433P	Guanine est remplacée par thymine en position 428	Transversion	Remplacement de sérine (S) en proline (P) à la position 1433
P472L	Cytosine est remplacée par thymine en position 1415	Transition	Remplacement de proline (P) en leucine (L) à la position 472
T1001I	Cytosine est remplacée par thymine en position 3267	Transition	Remplacement de thréonine (T) en isoleucine (I) à la position 1001
I2230T	Thymine est remplacée par cytosine en position 6954	Transition	Remplacement d'isoleucine (I) en glutamate (E) à la position 2230
T58I	Cytosine est remplacée par thymine en position 173	Transition	Remplacement de thréonine (T) en isoleucine (I) à la position 58
M429I	Guanine est remplacée par thymine en position 17523	Transversion	Remplacement de méthionine (M) en isoleucine (I) à la position 429
P4619 L	Cytosine est remplacée par thymine en position 3002	Transition	Remplacement d'asparagine (N) en tyrosine (Y) à la position 501
P4715L	Cytosine est remplacée par thymine en position 14408	Transition	Remplacement de proline (P) en leucine (L) à la position 4715
A4489V	Cytosine est remplacée par thymine en position 13730	Transition	Remplacement d'alanine (A) en valine (V) à la position 4489
L5784A	Guanine est remplacée par adenine en position 17352	Transition	Remplacement de leucine (L) en alanine (A) à la position 5784
P2908I	Cytosine est remplacée par adenine en position 2722	Transversion	Remplacement de proline (P) en isoleucine (I) à la position 2908
M902I	Guanine est	Transition	Remplacement de

Résultats et Discussion

	remplacée par adenine en position 2706		méthionine (M) en isoleucine (I) à la position 902
A1708D	Cytosine est remplacée par adenine en position 5388	Transversion	Remplacement d'alanine (D) en arginine (R) à la position 1708
P309L	Cytosine est remplacée par thymine en position 926	Transition	Remplacement de proline (P) en leucine (L) à la position 309
P822L	Cytosine est remplacée par thymine en position 2465	Transition	Remplacement de proline (P) en leucine (L) à la position 822
D70G	Adenine est remplacée par guanine en position 209	Transition	Remplacement d'aspartate (D) en glycine (G) à la position 70
P2046L	Cytosine est remplacée par thymine en position 6137	Transition	Remplacement de proline (P) en leucine (L) à la position 2046
S370L	Cytosine est remplacée par thymine en position 1109	Transition	Remplacement de sérine (S) en leucine (L) à la position 370
L71F	Cytosine est remplacée par thymine en position 211	Transition	Remplacement de leucine (L) en phénylalanine(F) à la position 71
P323L	Cytosine est remplacée par thymine en position 968	Transition	Remplacement de proline (P) en leucine (L) à la position 323
K977 Q	Adenine est remplacée par cytosine en position 2989	Transversion	Remplacement de lysine (K) en glutamine (Q) à la position 977
R226K	Guanine est remplacée par adenine en position 677	Transition	Remplacement d'arginine (R) en lysine (K) à la position 226
Y655T	Adenine est remplacée par guanine en position 194	Transition	Remplacement de tyrosine (Y) en thréonine (T) à la position 655

Résultats et Discussion

	T3255I	Cytosine est remplacée par thymine en position 9764	Transition	Remplacement de thréonine (T) en isoleucine (I) à la position 3255
--	--------	---	------------	--

III.1.3.3. Arbre phylogénétique de Nucléocapside N

L'arbre phylogénétique (Figure 44) a été réalisé sur la base du gène Nucléocapside N de différents variants de Coronavirus qui ont touché plusieurs pays du monde avec une taille de 419 pb. Comme pour l'arbre du gène S, et gène ORF1ab l'arbre du gène N est constitué de plusieurs variants, le variant Anglais (Alpha B.1.1.7), le variant indien (Delta B.1.617.2), le variant Brésilien (Gamma P.1), le variant Sud-africain (Omicron B.1.1.529), qui représentent des mutations apparues dans plusieurs pays.

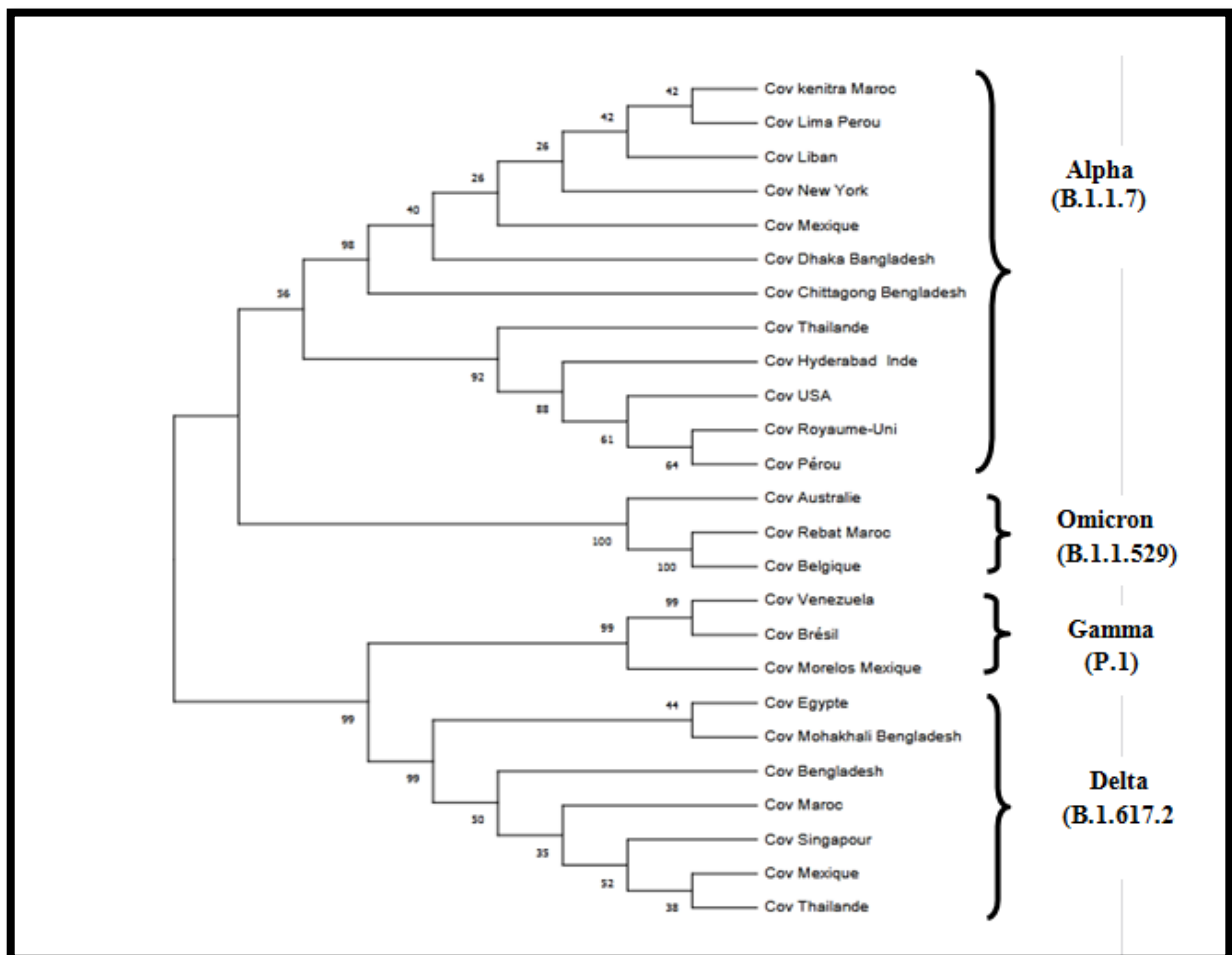


Figure 44 : Arbre phylogénétique de nucléocapside N représentant les séquences de différents variants de Coronavirus.

Résultats et Discussion

Les séquences de différents variants présentées dans l'arbre du gène de Nucléocapside N sont :

-Alpha (B.1.1.7) : la taille de la séquence Kenitra Maroc varie entre 28230-29456, Lima Pérou 28220 - 29479 pb, Liban 28440-29203 pb, New York 28205-29464 pb, Mexique (28217-29476pb, Dhaka Bengladesh 28245-29504pb, Thaïlande 28235 - 29494 pb, Hyderabad Inde 28258 - 29517pb, USA 28220-29479 pb, Royaume-Uni 28274 -29533pb, Pérou (28085-29344pb, ces séquences ont subi les mutations suivantes : Kenitra Maroc (S235F), Lima Pérou (S235P), Liban (P314L), New York (S235F, R203K, G204R), Mexique (G204R), Dhaka Bengladesh (Y73C), Thaïlande (L139F), Hyderabad Inde (P344S), USA (P13L D3L), Royaume-Uni (L37F, N1410T, A116V), Pérou (P344S).

-Delta (B.1.617.2) : la taille de la séquence égyptienne varie entre 28245-29504pb, Mohakhali Bengladesh 28268 - 29527 pb, Bengladesh 28207- 29466 pb, Maroc 28249-29508pb, Singapour 28220-29321, Mexique 28223- 29482pb, Thaïlande 28364-29546 pb. Elles comptent les mutations suivantes : Egypte (D63G, P344S), Mohakhali Bengladesh (A211V, T141I, G204R), Bengladesh (R203k), Maroc (D377Y), Singapour (P80R), Mexique (G215C), la Thaïlande (D63G).

-Gamma (P.1) : la taille de la séquence de Venezuela varie entre 228 - 21508 pb, Brésil 272 - 21546 pb, Morelos Mexique 264-21553 pb. Elles ont eu comme mutations: Venezuela (R203K K204R), Brésil (P80R, P344S), Morelos Mexique (R203K, G204R).

- Omicron (B.1.1.529) : la taille de la séquence d'Australie 230- 21556 pb, Rabat Maroc 241 - 21446 Pb, Belgique 212 - 13417 pb. Ces séquences contiennent les mutations suivantes : Australie (D63G, R203K G215C, D377Y), Rabat Maroc (R203K, G204R), Belgique (P240S). La protéine N compte 20 mutations dans le tableau 7, réparties sur la totalité de son code génétique, parmi ces mutations nous avons : G204R, R203K, P344S. D'après **Tobias et al., (2021)** ces mutations de la protéine N sont un modificateur majeur des interactions virus-hôte, augmentent la charge virale et confirment le potentiel de la protéine N, dans cette protéine ces mutations peuvent améliorer sa fonction en déclenchant la surexpression des gènes qui contribuent à l'exacerbation de la pathogenèse du SARS-COV -2 (**Tobias et al.,2021**).

Tableau 7 : Caractéristiques des différentes mutations que nous avons obtenues dans l'arbre du gène de la nucléocapside N.

Résultats et Discussion

Gène	Mutations	Base changée et sa position dans le génome	Type de mutation	Impact de la mutation sur la protéine
N	G204R	Guanine est remplacée par cytosine en position 610	Transversion	Remplacement de Glycine(G) en arginine (R) à la position 204
	L139F	Adénine est remplacée par cytosine en position 417	Transversion	Remplacement de leucine (L) en phénylalanine (F) à la position 139
	P314L	Cytosine est remplacée par thymine en position 15324	Transition	Remplacement de proline (P) en leucine (L) à la position 314
	R203K	Guanine est remplacée par adénine en position 608	Transition	Remplacement d'arginine (R) en lysine (K) à la position 203
	D3L	Adénine est remplacée par guanine en position 123	Transition	Remplacement de l'acide aspartique (D) en Leucine (L) à la position 3
	S235F	Cytosine est remplacée par thymine en position 704	Transition	Remplacement de sérine (S) en phénylalanine (F) à la position 235
	P344S	Cytosine est remplacée par thymine en position 1030	Transition	Remplacement de proline (P) en sérine (S) à la position 344
	L37F	Guanine est remplacée par thymine en position 111	Transversion	Remplacement de leucine (L) en phénylalanine (F) à la position 37
	N1410T	Thymine est remplacée par guanine en position 1233	Transversion	Remplacement d'asparagine (N) en thréonine (T) à la position 1410
	A116V	Cytosine est remplacée par thymine en position 347	Transition	Remplacement d'alanine (A) en valine (V) à la position 116
	P13L	Cytosine est remplacée par thymine en position 38	Transversion	Remplacement de proline (P) en leucine (L) à la position 13

Résultats et Discussion

G215C	Guanine est remplacée par thymine en position 643	Transversion	Remplacement de glycine (G) en cystéine (C) à la position 215
D377Y	Guanine est remplacée par thymine en position 1129	Transversion	Remplacement d'acide aspartique(D) en tyrosine (y) à la position 377
D63G	Adenine est remplacée par guanine en position 188	Transition	Remplacement d'acide aspartique (D) en glycine (G) à la position 63
A211V	Cytosine est remplacée par thymine en position 632	Transition	Remplacement d'alanine (A) en valine (V) à la position 211
T141I	Cytosine est remplacée par thymine en position 422	Transition	Remplacement de thréonine (T) en isoleucine (I) à la position 141
P80R	Cytosine est remplacée par guanine en position 239	Transversion	Remplacement de proline (P) en arginine (R) à la position 80
P240S	Cytosine est remplacée par thymine en position 718	Transition	Remplacement de proline (P) en sérine (S) à la position 240

III.1.3.4. Arbre phylogénétique de gène ORF3a

L'arbre phylogénétique (Figure 45) a été réalisé sur la base des séquences du gène ORF3a, dont la taille varie entre 25393 et 26220pb, de différents variants de Coronavirus qui ont touché plusieurs pays du monde. Le gène ORF3a est constitué de plusieurs variants, le variant Anglais (Alpha B.1.1.7), les variants sud-africain (Béta B.1.351) et (Omicron B.1.1.529), le variant indien (Delta B.1.617.2) et le variant Brésilien (Gamma P.1) qui entraînent une augmentation de la transmissibilité du virus et de la sévérité de la maladie (augmentation des hospitalisations ou des décès)(Konings *et al.*,2021).

Résultats et Discussion

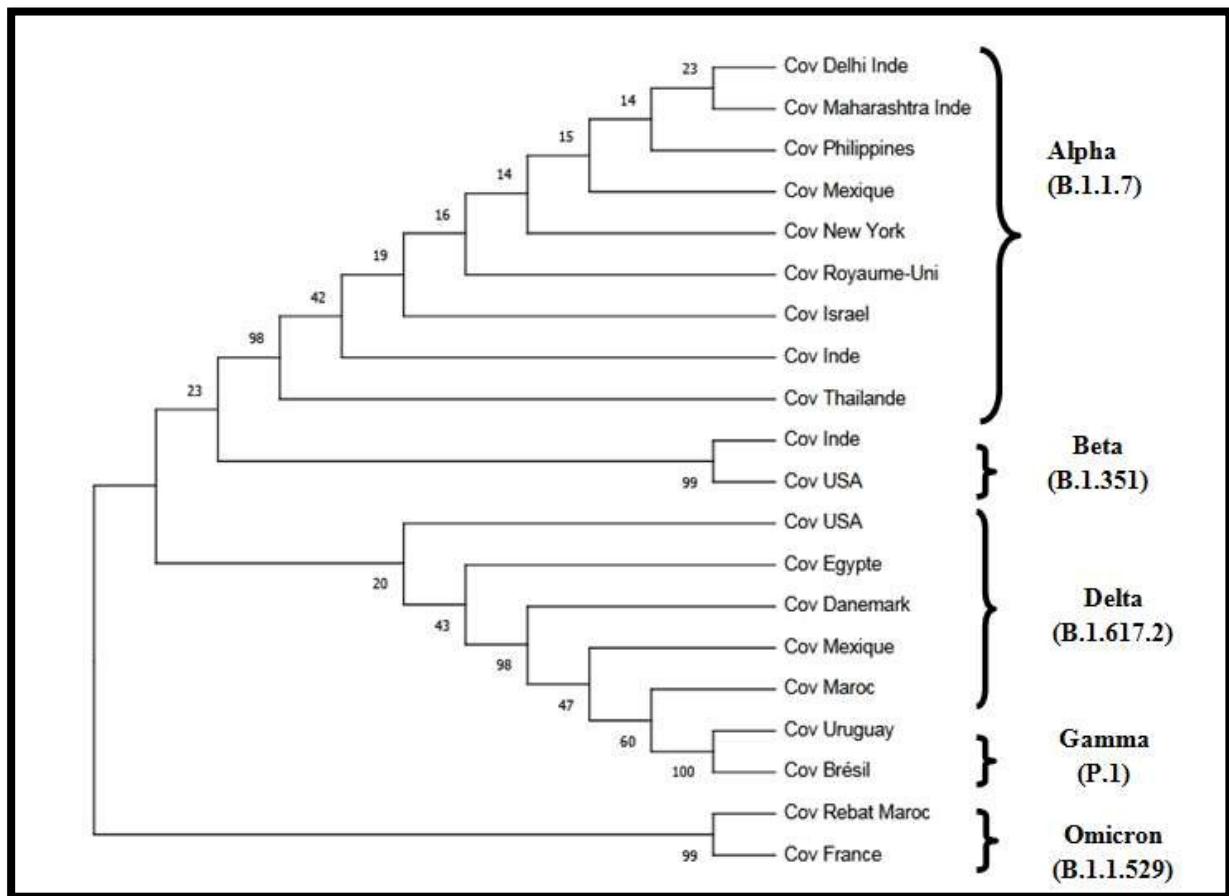


Figure 45 : Arbre phylogénétique de gène ORF3a représentant les séquences de différents variants de Coronavirus.

Les séquences de différents variants présentés dans l'arbre du gène ORF3a sont :

-**Alpha(B.1.1.7)** : les tailles des séquences sont comme suit : la taille de la séquence Delhi Inden25220-26456 pb, Maharachtra Inde 25336-26122 pb, Philippines 25355 - 26182pb, Mexique 25432-26222 pb, New York 25324 - 26151 pb, Royaume-Uni 25393 - 26220 pb, Israël 25391- 26218 pb, Inde 25364 - 26191pb, Thaïlande 25361 - 26188pb. Ces séquences ont les mutations suivantes : Delhi Inde (W128L, L140V G251V), Maharachtra Inde (S26L), Philippines (S171L, G251V, S26L), Mexique (L275F), New York (I82T), Royaume-Uni (G251V), Israël (Q57H, W131CW193L, T223I), Inde (Q57H), Thaïlande (F15L).

- **Beta (B.1.351)**: la taille de la séquence indienne varie entre 25393-26220 pb et USA 25141-25968 pb. Ces séquences ont différentes mutations tels que : Inde (Q57H) et USA (S171L, Q57H, S26L, G251V).

-**Delta (B.1.617.2)**: la taille de la séquence USA varie entre 25339- 26166 pb, Egypte 25364 - 26191 pb, Danemark (25362- 26189 pb), Mexique 25346 - 26173 p'Maroc 25368

Résultats et Discussion

- 26195 pb. Elles ont comme mutations : USA (Q57H), Egypte (Q57H, S26L), Danemark (Q57H, G251V), Mexique (S26L), Maroc (S26L),

-Gamma (P.1): la taille de la séquence d'Uruguay varie entre 25391 - 26218 pb, Brésil 25384 - 26211 pb. Différentes mutations sont détectées dans ces séquences qui sont : Uruguay (S253P, Q57H) et Brésil (S253P, S26L, G251V)

-Omicron (B.1.1.529): la taille de la séquence de Rabat Maroc varie entre 25347 - 26174 pb et la France de 25293 - 26120pb. Ces séquences présentent les mutations suivantes : Maroc (T64T, Q57H, G251V) et France (S92L, D155Y, S26L)

La protéine ORF3a compte 20 mutations dans le tableau 8, réparties sur la totalité de son code génétique, parmi ces mutations celles situées dans le domaine ORF3 (Q57H, S26L, G251V). Ces mutations ont un impact sur la gravité de la maladie ou l'efficacité du vaccin. Il a cependant été spéculé qu'elles conduisent à des concentrations plus élevées dans les voies aériennes supérieures et augmentent ainsi le taux de transmission du virus d'environ 50% (Shubham *et al.*, 2022).

Tableau 8 : Caractéristiques des différentes mutations des séquences que nous avons présentées dans l'arbre du gène ORF3a.

Gène	Mutations	Base changée et sa position dans le génome	Type de mutation	Impact de la mutation sur la protéine
ORF3a	F15L	Thymine est remplacée par Adénine en position 45	Transversion	Remplacement de phénylalanine (F) en leucine (L) à la position 15
	Q57H	Adénine est remplacée par thymine en position 171	Transversion	Remplacement de glutamine (Q) en histidine (H) à la position 57
	G251V	Guanine est remplacée par thymine en position 752	Transversion	Remplacement de glycine (G) en valine (v) à la position 251
	S26L	Cytosine est remplacée par thymine en position 77	Transition	Remplacement de sérine (S) en leucine (L) à la position 26
	I82T	Thymine est remplacée par cytosine en position 245	Transition	Remplacement d'isoleucine (I) en thréonine (T) à la position 82

Résultats et Discussion

W128L	Guanine est remplacée par thymine en position 383	Transversion	Remplacement de tryptophane (W) en leucine (L) à la position 128
L140V	Cytosine est remplacée par guanine en position 418	Transversion	Remplacement de leucine (L) en valine (V) à la position 140
S171L	Cytosine est remplacée par thymine en position 512	Transition	Remplacement de serine (S) en leucine (L) à la position 171
W131C	Guanine est remplacée par cytosine en position 393	Transversion	Remplacement de tryptophane (W) en cystéine (C) à la position 131
T223I	Cytosine est remplacée par thymine en position 668	Transition	Remplacement de thréonine (T) en isoleucine(I) à la position 223
L275F	Cytosine est remplacée par thymine en position 823	Transition	Remplacement de leucine (L) en phénylalanine (F) à la position 275
S171L	Cytosine est remplacée par Thymine en position 512	Transition	Remplacement de serine (S) en leucine (L) à la position 171
S253P	Thymine est remplacée par Cytosine en position 757	Transversion	Remplacement de serine (S) en proline (P) à la position 235
S92L	Cytosine est remplacée par thymine en position 275	Transition	Remplacement de serine (S) en leucine (L) à la position 92
D155Y	Guanine est remplacée par thymine en position 463G	Transversion	Remplacement d'acide aspartique (D) en tyrosine (Y) à la position 155

III.2. Discussion

III.2.1. Comparaison des bases azotées de gènes plus variable pour le SRAS - COV-2

D'après **Graw et al. (2006)**, plus l'ADN est riche en paires G/C plus il résiste à la dénaturation par l'augmentation de la température, le taux de G+C d'une molécule d'ADN est la fréquence relative, exprimé généralement en pourcentage. Les segments d'ADN composés de plusieurs bases G-C sont plus stables que les séquences composées de plusieurs bases A-T, donc les gènes qui ont un taux de CG élevé subissent moins de mutation que les autres car ils sont plus conservés. Selon les travaux que nous avons réalisés sur l'analyse de la variabilité des gènes les plus variables du SARS-CoV-2 et les mutations qui les touchent, nous avons remarqué que le gène S est le gène le plus variable car le taux de mutation dans ce gène est très élevé en comparant avec la variabilité génétique des gènes N, ORF1ab et ORF3a. Cette variabilité est due aux pourcentages différents des bases azotées, chose qui a été prouvée par **Xia et Xie (2001 ; 2013)** qui expliquent que les variations des bases signifie qu'il y a des mutations. De plus, dans une étude faite à l'aide du programme DnBA (DNA Basis Analysis), par **Derouiche et al. (2021)**, ou ils ont trouvé que le pourcentage de AT dans les gènes les plus variables est plus élevé par rapport à celui de GC commençant par le gène S (A :29.435%, T :33.333%, C :18.864%, G :18.367%), le gène ORF1ab (A :30.089%, T :32.300%, C :17.703%, G :19.900%), le gène ORF3a (A :27.174%, T :33.454%, C :20.014%, G :18.357%) et le gène N (A :31.984%, T :20.952%, C :24.841%, G :21.825%). D'après cela et selon nos résultats, nous pouvons classer ces gènes par ordre de variabilité qui est traduite par le taux de mutation, premièrement le gène S, deuxièmement le gène ORF1ab, troisièmement le gène ORF3a et enfin le gène N. Nos résultats sur les mutations reflètent la variabilité de ces gènes qui a une relation avec le pourcentage des bases A, T, C, G sachant que ces gènes sont moins riches en GC par rapport à AT, et donc effectivement ces gènes sont instables et subissent plus de mutations par rapport au reste des gènes du SARS-CoV-2 à cause du taux élevé des bases A et T.

III.2.2. Comparaison entre les séquences des variants de différents pays dans le monde

Les arbres phylogénétiques des gènes S, ORF1ab, N et ORF3a se composent de plusieurs clades représentant les variants du SARS-CoV 2 de différents pays par continent qui sont : variant Anglais (Alpha B.1.1.7), variants sud-africain (Béta B.1.351) et (Omicron B.1.1.529), variant Brésilien (Gamma P.1), et le variant indien (Delta B.1.617.2) (**Konings et al., 2021**).

Résultats et Discussion

Nous pouvons classer les pays infectés par les différents variants du SARS-CoV 2 selon des catégories ; la première catégorie représente les pays où il y a une conservation des mêmes mutations par exemple dans les arbres des quatre gènes, nous constatons que les séquences de tous les variants de l'Afrique sont regroupées avec des séquences de différents pays du monde partageant ainsi les mêmes mutations, cette similarité génétique peut être due au facteur économique /commercial qui semble être la raison la plus possible pouvant justifier la phase initiale de diffusion du Covid19. Selon **Bontempi et al. (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 les moyens de transport. De plus, l'apparition de la séquence du variant omicron de l'Afrique du Sud regroupé dans l'arbre phylogénétique qui représente le gène S (figure 42) avec Bangladesh, Australie, Maroc, Chine et Belgique peut être traduite par la similarité des mutations de la protéine spike dont les substitutions de l'acide aminé N (Asparagine) par d'autres acides aminés dans des positions proches ; nous avons :N440K, N501Y, N211I, N764K, et N856K (Tableau 3). Cela a coïncidé avec une accélération de l'incidence de cas de Covid19 et le variant omicron a rapidement été détecté dans les pays avec lesquels elle est regroupée. Cette similarité entre ces pays qui sont géographiquement éloignés explique qu'il y a des facteurs et des conditions environnementales tels que : température et humidité, transmission par les aliments et leurs emballages, l'air, la population particulièrement jeune, seraient à l'origine d'une propagation rapide de l'infection (**Hoummadi et al., 2020**). Nous remarquons aussi que dans les arbres phylogénétiques des gènes S, ORF1ab, N et ORF3a représentés dans les figures 42, 43,44 et 45 respectivement, dans les clades qui regroupent les séquences du variant delta nous trouvons toujours le Maroc et l'Egypte dans des positions très proches c'est-à-dire ils partagent la même séquence avec les mêmes mutations, cela est justifié par les travaux de **Laamrti et al. (2020)** qui a trouvé que toutes les variants de SARS Cov 2 d'Afrique du Nord portant la même mutation D614G localisée sur le gène S. Aussi dans l'arbre du gène ORF1ab représenté dans la figure 43, nous avons constaté que les séquences de la souche bêta de la Colombie et la Chine sont localisées dans le même clade, chose qui a été prouvée par **Chu et al.(2021)** qui a montré que la quatrième vague épidémique de Covid19 à Hong Kong a été introduite à partir de la Colombie par un voyageur de Colombie à Hong Kong.

La deuxième catégorie regroupe les pays qui ont des mutations différentes au sein d'un même pays. La transmission mondiale de COVID-19 offre au SRAS CoV-2 une opportunité abondante de subir des mutations rapides. Un autre mécanisme important pour l'évolution du SARS-CoV-2 est la sélection naturelle, qui rend le virus plus infectieux tout en étant moins

Résultats et Discussion

virulent, en général (Nathan et al., 2020). C'est pour cela que nous pouvons trouver des mutations dans les séquences d'un même pays comme le cas dans l'Inde et l'USA. Nous avons constaté que les deux séquences alpha de l'Inde représentées dans l'arbre du gène ORF1ab (figure 43) sont éloignées ce qui montre clairement qu'il n'y a pas une seule et même mutation qui a touché les deux séquences et d'après les résultats de Jahvani et al.(2022), le variant alpha du virus était prédominant dans le nord de l'Inde par rapport au sud ainsi que sa présence modérée dans le sud était en raison de la présence du variant delta qui est plus virulent et qui a finalement pris le paysage complet de l'Inde après ça. De plus, Kaushal et al. (2020) a montré que la majorité des gènes des variants qui ont touché les USA accumulaient des mutations. C'est pour cela, en comparant l'apparition des différentes séquences de l'USA représentées dans tous les arbres, nous voyons que les positions diffèrent pour chaque gène.

II.2.3. Les différentes mutations du Coronavirus

Les variants présentés dans nos arbres phylogéniques dont nous avons parlé précédemment sont dus à des mutations qui ont touchées les différents gènes du Coronavirus.

III.2.3.1. Les mutations du gène S

D'après Fouratiet al. (2020) les variants identifiés qui augmentent le risque pour la santé humaine sont : Alpha, Beta, Delta, Gamma, Omicron présentant des mutations génétiques de la protéine spike ; il a été démontré une augmentation de la transmissibilité ou un impact sur l'épidémiologie de SARS-COV-2. Il existe plusieurs mutations qui ont un impact très important sur la protéine de la pointe, parmi ces mutations nous avons :

La délétion H69/V70 : présente dans deux variants qui sont Alpha(B.1.1.7) et Omicron (B.1.1.529), ce qui a donné plusieurs sous-clades de Alpha comme : Allemagne, Philippines, Mexique. Cette mutation augmente également dans le variant Omicron la transmissibilité et diminue la neutralisation du virus par les anticorps (Kemp et al., 2020).

En Allemagne, d'après une étude réalisée par des chercheurs japonais Takano et al., (2021) la délétion H69/V70 à elle seule entraîne un prolongement de l'inflammation (dure plus longtemps), une augmentation de la charge virale ainsi qu'une aggravation de la maladie.

La mutation K417T : elle est impliquée dans des changements structurels de la protéine S et l'évasion immunitaire. Elle est présente dans le groupe Alpha (B.1.1.7), Beta (B.1.351) et Delta (B.1.617.2), d'où vient la division de notre arbre en sous-clades de plusieurs pays tels que: Maroc, Inde, USA, Suisse.

Aux USA, cette mutation semble aussi diminuer l'affinité de la protéine S pour les récepteurs ACE2 qui d'après de récentes études semble être due à la rupture de la liaison

Résultats et Discussion

hydrogène entre la lysine 417 et l'acide aspartique 101 provoquée par la mutation N417, cependant cette diminution est compensée par la présence de la mutation N510Y (**Boehm et al., 2021**).

La mutation L452R : que l'on retrouve dans les trois clades Alpha, Delta et Gamma. Cette mutation est impliquée dans l'augmentation de la transmissibilité et l'évasion immunitaire (**Kannan et al., 2021**).

La mutation T478K: présente dans le groupe Delta (B.1.617.2) et le sépare en 6 sous clades qui représentent les pays comme : Maroc, Inde, Afrique du sud. Selon **Saathvik et al.(2021)**elle entraîne une diminution de la neutralisation du virus par les anticorps monoclonaux, les sérums de convalescence et les sérums post-vaccination. La mutation T478K perturbe la fixation des anticorps monoclonaux sur la protéine S qui est due à la différence dans la longueur des chaînes latérales entre la lysine (plus longue) et la thréonine.

La mutation E484K : Cette mutation modifie la forme de l'extrémité de la protéine S entraînant une augmentation de l'affinité pour l'ACE2.Présente dans le groupe Alpha (B.1.1.7), Beta (B.1.351) et Delta (B.1.617.2) et distribuée dans plusieurs pays dans notre arbre tels que : Allemagne, Inde, Danemark. La mutation E484K permet aussi d'échapper partiellement aux anticorps principalement les anticorps polyclonaux issus d'une infection ou vaccination antérieure (**Larsen et al., 2020**).

En Inde, la mutation E484K augmente sensiblement la quantité d'anticorps sériques nécessaires pour empêcher l'infection des cellules (**Wise et al., 2021**).

Au Danemark, selon **Grenay et al.(2021)**la mutation E484K semble réduire significativement la neutralisation du virus par le sérum des convalescents. Dans une étude menée par **Nonaka et al. (2021)** une femme précédemment infectée par un variant sans la mutation E484Q a été réinfecté par un variant présentant la mutation E484K.

En Allemagne, selon une étude publiée par **Liu et al. (2021)** la mutation E484K entraîne la réduction de la neutralisation du virus par les anticorps monoclonaux. La mutation E484K augmente la liaison à l'ACE2 d'autant plus en présence de K417N et N501Y (**Zahradnik et al., 2021**), augmentation in silico. C'est une mutation majeure entraînant la réduction de la fixation des anticorps sur RBD et de la capacité de neutralisation de sérums de convalescents à 13 mois datant de février 2020 (**Greaney et al., 2020**), ou de plasma de sujet convalescent

Résultats et Discussion

de COVID-19 et entraînant l'échappement à la neutralisation (**Andreano et al., 2020**) ou une diminution nette (**Jangra et al., 2020**).

La mutation N501Y : présente dans les clades Alpha (B.1.1.7) et Omicron (B.1.1.529) donnant plusieurs sous clades des différents pays présentés dans notre arbre. Cette mutation modifie légèrement l'extrémité de la protéine S ce qui augmente l'affinité de la protéine S pour les récepteurs ACE2 améliorant ainsi l'attachement virale et l'entrée du virus dans les cellules hôtes par conséquent une augmentation de la transmission (**Larsen et al., 2020**).

Dans l'Afrique du Sud, d'après une étude récente effectuée dans le centre de recherche de **Thomas et al. (2020)** des chercheurs ont proposés une explication sur l'augmentation de la transmissibilité des variants portant la mutation N501Y. cette recherche suggère que la nature hydrophobe de la protéine Y501 (tyrosine) se lie plus étroitement aux fragments hydrophobes situés dans les récepteurs ACE2 augmentant ainsi l'affinité entre la protéine S et les récepteurs ACE2 et favorisant l'évasion aux anticorps, par contre la protéine N501 est hydrophile (**Luan et al., 2021**).

D'après la recherche publiée par **Nelson et al. (2021)** les mutations K417N, E484K et N501Y augmentent l'affinité de la protéine S pour les récepteurs ACE2, l'affinité est d'autant plus augmentée lorsque les deux mutations (E484K et N501Y) sont présentes dans le même variant. La mutation K417N semble stabiliser la mutation N501Y et ainsi stabiliser la liaison protéine S – récepteur ACE2.

Les mutations K417N/T, L452R, E484K, E484Q et N501Y sont situées au niveau du domaine RBD qui est constitué de 71 acides aminés situé entre la position 438 et 508, et est ainsi un point clé pour les mutations qui peuvent avoir une grande incidence clinique en termes de virulence, de transmissibilité et d'évasion immunitaire de l'hôte (**Nelson et al., 2021**).

La mutation D614G : Cette mutation est située en dehors du site RBD et donc ne modifie pas l'affinité de la liaison avec les récepteurs ACE2 ou la neutralisation du virus par les anticorps, cependant elle peut augmenter la densité de la protéine S en préservant son intégrité structurale et évite la perte de la sous unité S1 entraînant en conséquence une augmentation de la transmissibilité du virus ce qui explique le fait qu'elle soit la mutation la plus répandue dans le monde (actuellement détectée dans presque tous les variants en circulation) et qu'elle est remplacé par la souche ancestrale (**Zhang et al., 2020**). Cette mutation est présente dans plusieurs pays que nous avons représenté dans l'arbre dans le groupe Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), Gamma (P.1), les pays sont : Maroc, Inde, Venezuela, Suisse, Malaisie.

Résultats et Discussion

En Malaisie, la glycine entraîne un changement de conformation avec augmentation de l'état « up » du RBD et amélioration du clivage par la furine dans SD2 des sous-unités S1 et S2 de la forme de Spike pré fusionnelle. (Flores et al., 2020)

En Suisse, la mutation D614G augmente la densité de la protéine S à la surface des particules virales et la liaison au récepteur ACE2 (Zhang et al., 2020, Benvenuto et al., 2020), ainsi que l'augmentation de la transmissibilité des souches liées à une charge virale plus élevée lors de l'infection par le mutant G614.

III.2.3.2. Les mutations du gène ORF1ab

Notre arbre phylogénétique se compose de cinq clades principaux qui sont : Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), Gamma (P.1), et Omicron (B.1.1.529). Dans le gène ORF1ab, 4 principales mutations ont été observées dans les séquences de différents variants du SARS COV-2(I2230T, P314L, T1001I, P4619L). Ces mutations sont dominantes dans la très grande majorité des pays du monde (90%) ; elles entraînent une augmentation de la transmissibilité des souches (Cheng et al.,2020).

La mutation I2230T : est présente dans tous les clades constituant l'arbre et le divisant en plusieurs sous clades présentés par les pays suivants : Mexique, Inde, New York, la Colombie, le Maroc, l'Egypte, elle est majoritaire dans tous les pays du monde. Cette mutation entraînerait une augmentation de la transmissibilité des souches (Cheng et al., 2020). Au Mexique, en raison de cette mutation I2230T, la population mexicaine a une prévalence élevée de SARS-CoV-2 par rapport aux autres pays, de sorte que la description de la séquence virale de ces minorités est importante pour comprendre la transmission du virus et la maladie (David et al., 2021).

En Inde, la mutation I2230T est associée à un risque accru de transmission du virus selon des données publiées dans des revues scientifiques à comité de lecture, cette mutation est dangereuse car le nombre de cas en Inde a fortement augmenté, et la prévalence de la variante a également conduit à une augmentation du nombre de cas (Shubham et al., 2022).

A New York, d'après une équipe de chercheurs Caserta et al.(2021) un nouveau type de Coronavirus a été découvert qui se propage très rapidement et porte la mutation I2230T qui nuit à l'efficacité des vaccins. Les patients infectés par le virus porteur de la mutation I2230T sont susceptibles d'être hospitalisés.

La mutation P314L : est une mutation importante du gène ORF1ab qui est apparue seulement sur les variants suivantes : Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2) et Omicron(B.1.1.529) de notre arbre. Cette mutation dangereuse montre la distribution

Résultats et Discussion

géographique locale dans des pays tels que la Thaïlande, le Mexique, Israël et la Chine. Selon **Korber et al. (2020)** une mutation grave se propage facilement dans la population en raison de sa grande capacité de mémorisation.

En Thaïlande, d'après **Phutthasopit et al. (2021)** l'exception la plus notable est la mutation P314L, dont la prévalence augmente considérablement l'infection à l'échelle mondiale.

En Chine, la mutation du gène ORF1ab (P314L) est apparue et a rapidement dépassé les autres en termes de propagation, y compris l'isolat original du SRAS-2 de Wuhan, en Chine. Cette variante est une caractéristique de la branche la plus répandue (A2a) des génomes du SARS-CoV-2 dans le monde (**Bhattacharyya et al., 2020**). En utilisant des données évolutives, plusieurs groupes ont suggéré que la mutation P314L pourrait conférer une transmissibilité accrue conduisant à une sélection positive, tandis que d'autres ont fait valoir que les preuves actuellement disponibles ne soutiennent pas la sélection positive (**Dorp et al., 2020**).

Au Mexique, la mutation dans le domaine ORF1ab représente un facteur d'attraction lentement indépendant de l'espace et du temps, et pourrait servir de marqueur moléculaire pour suivre l'état du virus dans divers endroits au Mexique. Cette mutation est importante et a un impact significatif sur l'homme et est plus dangereuse (**Taboada et al., 2020**).

La mutation T1001I : cette mutation entraîne une augmentation de la transmissibilité des cinq variantes de notre arbre dans les différents pays qui représentent les sous clades de tous les variants dont : Mexique, Inde, Bangladesh et Colombie.

Au Mexique, la mutation T1001I a montré des risques potentiels. De plus, elle est un type de mutation important parmi d'autres types de mutations du SRAS, permettant une augmentation de la densité protéique à la surface des particules virales CoV2 (**David et al., 2021**).

En Colombie, cette mutation est importante pour sa susceptibilité à l'immunité de l'hôte et la performance des tests de diagnostic et de certains vaccins. Certaines des races importantes qui ont cette mutation sont : Alpha, Beta, Gamma et Omicron. C'est une mutation importante pour comprendre les variants génétiques émergentes et ce qu'elles ont en corrélation avec la gravité de la maladie (**Jubby et al., 2021**).

Au Venezuela, l'effet de cette mutation T1001I est d'échapper aux anticorps neutralisants et augmenter leur fréquence, l'observation les a amenés à soupçonner que les réinfections pourraient être associées à l'existence de cette mutation. Contrairement à leurs attentes, les cas documentés de réinfection dans le monde n'ont pas été liés à la présence de mutations (**Loureiro et al., 2022**).

Résultats et Discussion

La mutation P4619L : cette mutation est détectée dans différents pays qui représentent les sous clades des cinq variants principaux de notre arbre comme : Bangladesh, Inde, Chine, Belgique. Cette dernière a été sélectionnée car elle est très probablement associée à une fuite immunitaire et à une transmissibilité accrue. Une combinaison de cette mutation peut permettre de soupçonner certaines variantes, mais un séquençage peut être nécessaire pour le confirmer. Étant donné que les résultats de la PCR sont présentés dans un délai plus court que le séquençage, des mesures spécifiques peuvent être prises une fois que des cas porteurs de cette mutation sont détectés (**Boccia et al., 2022**).

III.2.3.3. Les mutations du gène N

Notre arbre phylogénétique se compose de quatre groupes qui sont : Alpha (B.1.1.7), Delta (B.1.617.2), Gamma (P.1), Omicron (B.1.1.529). Dans la protéine de Nucléocapside N, nous avons analysé 3 principales mutations (G204R, R203K, P344S). Plusieurs études expérimentales ont montré que ces mutations contribuent à l'amélioration de l'infection virale, car elles ont un effet sur la flexibilité de la structure protéique lors des simulations, modifient les fonctions et réduisent l'aptitude du coronavirus, et ont des effets négatifs sur la fonction protéique (**Tobias et al., 2021**).

Les mutations G204R/R203K: deux mutations de la protéine N qui pourraient être responsables d'une plus grande transmissibilité des quatre groupes (Alpha B.1.1.7, Delta B.1.617.2, Gamma P.1, Omicron B.1.1.529) qui sont apparues dans différents endroits dans le monde tel que : Mexique, New York, Bangladesh, Belgique, Pérou, Maroc.

Au Mexique les deux mutations 203K/204R ont montré une augmentation de l'infection dans la lignée cellulaire pulmonaire humaine et ont entraîné une augmentation des dommages aux vaisseaux sanguins dans les poumons du patient. Ils ont observé une association positive entre la sévérité accrue de COVID-19 et la fréquence d'occurrence de 203K/204R (**Taboada et al., 2022**).

Au Bangladesh, ces mutations 203K/204R ont contribué à l'augmentation de la transmission et de la virulence du SARS-CoV-2 qui sont importantes pour la propagation du virus lors d'une pandémie (**Haibo et al., 2021**).

A New York, l'effet de ces mutations a entraîné une augmentation marquée des cas, les cas documentés d'infection dans le monde n'étaient pas associés à la présence de mutations 203K/204R. Même pendant la période de mars à mai 2021, lorsque sa fréquence a augmenté et s'est considérablement étendue (**Caserta et al., 2021**).

Résultats et Discussion

La mutation P344S: a été détecté au Brésil, Egypte et en Inde avec une transmissibilité rapide cette mutation augmente la densité de la protéine N, ainsi que l'augmentation de la transmissibilité des souches liée à une charge virale. Cette mutation était répandue dans le monde entier et fréquente dans la protéine N (**Khan et al., 2020**).

III.2.3.4. Les mutations du gène ORF3a

L'arbre phylogénétique du gène ORF3a est constitué de cinq groupes qui sont: Alpha(B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), Gamma (P.1) et Omicron (B.1.1.529). Dans ce gène, 3 principales mutations ont été observées dans les séquences de différents variants du SARS-CoV-2 (Q57H, S26L, G251V). D'après **Taboada et al., (2022)** ces mutations neutralisent une partie de la réponse immunitaire des personnes immunisées contre les lignées historiques du SARS-CoV-2 et conduisent à une augmentation rapide de la transmissibilité des souches de ce virus.

La mutation Q57H : est caractérisée par un remplacement de la glutamine (Q) en histidine (H) à la position 57 d'une protéine. Cette mutation est apparue sur les variants Alpha, Beta, Delta et Omicron ou était l'une des causes de l'apparition de plusieurs sous clades de ces derniers représentés par les différents pays du monde comme : l'Inde, Israël, USA, Egypte et Maroc.

En Inde, des études ont montré qu'une mutation Q57H dans la protéine ORF3a du SRAS-CoV-2 le rend plus infectieux et transmissible (**Shubham et al., 2022**).

En Egypte, il existe des preuves suggérant que la possibilité d'une mutation Q57H affectant l'efficacité du vaccin est très faible. Étant donné que la mutation ne se trouve pas dans le domaine de liaison au récepteur de la protéine de pointe, il est moins probable qu'elle affecte la capacité du domaine à induire des réponses immunitaires de l'hôte, ce qui est considéré comme une condition préalable à la neutralisation du virus médié par les anticorps (**Seadawy et al., 2020**)

Au Maroc, d'après **Melloulet al. (2021)** la mutation Q57H du gène ORF3a a augmenté les capacités d'infection, de répllication et de compétition ainsi que la transmissibilité du virus dans des cellules primaires humaines et dans un modèle humain.

La mutation S26L : d'après **Taboada et al. (2022)** est une mutation non synonyme au sein du domaine ORF3a mais elle est commune dans les cinq lignées du SARS-CoV-2 représentées dans l'arbre ; identifiée en Inde, Philippines, États-Unis et au Mexique. C'est l'un des principaux résidus de contact au sein d'ORF3a. Cette mutation affecte la sévérité de la maladie et l'efficacité du vaccin. Cependant, l'augmentation du taux de transmission du virus. L'étude démontre également que la mutation S26L a une apparence similaire à la souche

Résultats et Discussion

Parente et reste sensible à la neutralisation par des anticorps dirigés contre la forme primaire du SARS-CoV-2. Ces résultats sont donc encourageants pour la recherche de vaccins contre le virus.

La mutation G251V : elle est présente dans tous les variants existants dans l'arbre et distribuée par les sous clades qui représentent les pays suivants : USA, Danemark, Maroc, Brésil. Cette mutation affecte la reconnaissance des anticorps et permet l'échappement immunitaire du SARS-CoV-2. Il a été démontré que le virus porteur de cette mutation échappe à la reconnaissance des anticorps dans les sérums des personnes convalescentes et peut donc altérer l'efficacité des vaccins (**Taboada et al., 2022**).

En Danemark, d'après **Lassauniere et al. (2021)** la mutation G251V est devenue répandue dans ce pays, indiquant un avantage de forme par rapport à la souche originale de Wuhan qui permet une propagation plus rapide. Plusieurs études récentes indiquent que la mutation de G251V augmente l'interaction dans le domaine ORF3a, ce qui conduit à une augmentation de la protéine ORF3a totale intégrée au virus et donc à des particules virales plus stables.

III.2.3.5. Vaccin et mutations

Plus de 121 millions de cas confirmés de Covid ont été enregistrés et la maladie a fait plus de 2,7 millions de morts dans le monde, comme enregistré le 20 mars 2021 (**OMS, 2021**). Alors que la population mondiale s'adapte à ce qui est appelée la « nouvelle normalité » de la distanciation sociale, du port de masque et du contrôle de la température, une approche médicale efficace semble être la seule voie pour que la vie revienne à ce qu'elle était avant la pandémie. De nombreux pays ont utilisé en premier avec l'aide de la science et de la technologie de pointe où des scientifiques et des experts de la santé se sont réunis pour accélérer les traitements prophylactiques et curatifs potentiels afin de freiner la propagation et les effets néfastes du virus sous la forme de nouveaux traitements et vaccins efficaces.

Dans le présent travail, en étudiant les mutations de chaque gène nous avons constaté que le gène le plus muté est le gène de la protéine de la pointe S, cette protéine a attiré beaucoup d'attention en raison de son rôle critique dans l'infection virale et le développement de vaccins et d'anticorps. En parallèle, des études ont rapporté que la stabilité mécanique du SARS-CoV-2 RBD joue un rôle essentiel dans l'augmentation de la propagation du COVID-19 en raison d'interactions intermoléculaires plus fortes par rapport au RBD du SARS-COV (**Moreira et al., 2020**). Par conséquent, le suivi du taux de croissance des mutations existantes sur la protéine S RBD nous permet de surveiller les mutations qui peuvent avoir un impact sur

Résultats et Discussion

l'efficacité de médicaments anti-virus, d'anticorps et de vaccins, parmi ces mutations nous avons : N501Y, L452R, T478K, N501T, N550K, F490S, V483F, L452M et A348S qui sont communes à travers plusieurs pays au monde et qui ont des changements relativement élevés de la liaison de la protéine S et de l'ACE2, ce qui suggère qu'ils peuvent conduire à des variants plus infectieux.

Le 24 janvier 2020, le Centre chinois de contrôle et de prévention des maladies a réussi à isoler complètement la première souche chinoise de nouveau coronavirus (**WHO-China 2020**). À ce jour, la recherche et le développement mondiaux de vaccins contre le coronavirus ont couvert diverses catégories de vaccins, notamment les virus vivants et les vaccins inactivés, les vaccins sous-unitaires, les vaccins vectoriels, les vaccins à acide nucléique (vaccins à ARNm et vaccins à ADN), etc. (**Gao et al., 2020**). En mars 2021, treize vaccins avaient été approuvés pour demande alors que plus de 90 vaccins candidats sont en cours d'essais cliniques. Selon **Amanat et Krammer (2020)**, deux étapes importantes sont généralement nécessaires avant qu'un vaccin ne soit testé par des essais cliniques. Les vaccins sont d'abord testés pour leur capacité à fournir une immunité protectrice dans des modèles d'animaux appropriés, suivis du test de toxicité de ces vaccins sur des animaux tels que des souris (**Sharma et al., 2020**), des furets (**Shi et al., 2020**) et des singes (**Munoz et al., 2020**) pour garantir la sécurité d'application.

Les connaissances issues du développement du vaccin contre le SARS-CoV sont importantes et ont servi de base pour décrire le développement du SRAS -Vaccins CoV-2 (**Ma et al., 2020**). Des études ont montré que les vaccins ciblant la sous-unité RBD, S1 ou S2 du SRAS-CoV-2 présentent certains effets protecteurs contre le COVID-19 (**Zhang et al., 2020**). Ainsi, les nouveaux vaccins contre le coronavirus qui font actuellement l'objet de recherches sont généralement conçus et développés pour déstabiliser la protéine S et perturber ou affaiblir les interactions RBD.

III.3.Covid 19 en Algérie :

L'Algérie, comme le reste du monde a fait face à la propagation de ce virus représenté par plusieurs variants traduisant la présence de différentes mutations (**Zeghib et al., 2021**). Dans le cadre de l'activité de séquençage du virus SARS-CoV-2 mise en place par l'Institut Pasteur en Algérie et les labos de recherche pour la détection des variants circulants par dosage PCR, ils ont rapporté la détection des cinq variants : Alpha, Beta, Delta, Gamma et Omicron. Ces variants ont infecté plusieurs pays dans la même période dite vague que nous avons représentés dans les quatre arbres phylogénétiques obtenus pour le gène S (Figure 1), le gène ORF1ab (Figure 2), le gène N (Figure 3), et le gène ORF3a (Figure 4).

Première vague : Variant Alpha

Identifié pour la première fois par l'OMS au Royaume-Uni ; apparue en Algérie dans la première vague le 25 février 2020, le premier cas du variant Alpha est introduit par un Italien salarié d'Eni, originaire de Lombardie, l'une des zones les plus touchées en Italie, arrivé le 18 février, sur la base Menzel Ledjmet Est dans la wilaya de Ouargla. Le pic de cette vague a enregistré 453 décès et 4154 cas confirmés par PCR. Ce variant est caractérisé par plusieurs mutations du gène S (Tableau 3).) entre autres deux mutations majeures qui sont N501Y et D614G, identifiées dans notre arbre phylogénétique du gène S à Berlin en Allemagne (N501Y), Mexique, Allemagne et la Suisse (D614G). La mutation N501Y implique le remplacement par transversion d'Adénine par Thymines en position 23063 traduite par le remplacement de l'acide aminé Asparagine par l'acide aminé Tyrosine à la position 501. Elle modifie légèrement l'extrémité de la protéine S ce qui augmente son affinité pour les récepteurs ACE2 améliorant ainsi l'attachement virale et l'entrée du virus dans les cellules hôtes par conséquent une augmentation de la transmission (**Larsen et al., 2020**). La mutation D614G correspond à une substitution d'une Glycine par une Aspartique en position 614 dû à une transition d'Adénine par une Guanine en position 23403. Elle est située en dehors du site RBD et donc ne modifie pas l'affinité de la liaison avec les récepteurs ACE2 ou la neutralisation du virus par les anticorps, cependant elle peut augmenter la densité de la protéine S en préservant son intégrité structurelle et évite la perte de la sous-unité S1 entraînant en conséquence une augmentation de la transmissibilité du virus ce qui explique le fait qu'elle soit la mutation la plus répandue dans le monde (**Zhang et al., 2020**) .

Résultats et Discussion

; en parallèle les travaux de **Laamrti et al. (2020)** ont confirmé que toutes les variants de SARS-CoV-2 d'Afrique du Nord portant la même mutation D614G localisée sur le gène S.

Deuxième vague : Variant Beta

La deuxième vague est liée au variant Beta découvert pour la première fois en Afrique du Sud (**OMS, 2022**), apparu en Algérie en juin 2020 ; la première infection confirmée connue de Beta provient d'un échantillon prélevé le 9 juin. Selon l'Institut Pasteur, depuis le début jusqu'à la fin de cette vague le nombre total de cas est de 11300 avec 2985 décès. Ce nouveau variant est nommé Béta ou 501Y.V2 en raison de la mutation N501Y qu'il porte et qui est détectée dans pas mal de pays à travers le monde ; dans notre arbre du gène S il a été détecté en Afrique du sud. La mutation N501Y se situe dans la protéine spike à proximité du site de clivage, qui joue un rôle majeur dans la fusion entre les membranes virale et cellulaire (**Greaney et al., 2020**).

Troisième vague : Variant Delta

Dans la troisième vague le variant Indien Delta est apparu, l'Institut Pasteur d'Algérie a annoncé que les activités de séquençage mises en place pour surveiller les variants du SARS-CoV-2 ont permis de confirmer la présence de 7 premiers cas du variant Delta au niveau de la wilaya d'Alger depuis le 3 mai 2021. Avec ce variant l'Algérie a connu une forte augmentation du nombre de cas enregistrés ainsi que de mortalité, le nombre total des cas infectés confirmés par PCR a dépassé 14036 et 4504 décès. Ce variant est plus contagieux, plus virulent et plus transmissible par rapport au variant anglais, et moins sensible aux vaccins, il se caractérise par deux mutations spécifiques L452R et E484K touchant le gène S (**Larsen et al., 2020**). La mutation L452R a été détectée dans notre arbre du gène S dans les séquences d'ADN d'Inde (Delhi) et du Maroc ; elle correspond à une transversion d'une Thymine par Guanine en position 1355 traduite par une substitution d'un Thréonine en Arginine à la position 19. Cette mutation est impliquée dans l'augmentation de la transmissibilité et l'évasion immunitaire (**Kannan et al., 2021**). La mutation E484K présentée dans l'arbre du gène S identifiée dans la séquence prévenue du Danemark ; provoque le remplacement d'une Guanine par Adénine en position 1450 ce qui induit le remplacement d'acide aminé Glutamique par une Lysine en position 484.

Cette mutation permet d'échapper partiellement aux anticorps principalement les anticorps polyclonaux issus d'une infection ou vaccination antérieure (**Larsen et al., 2020**).

Résultats et Discussion

Quatrième vague : Variant Gamma

La quatrième vague est due au variant P1 (Gamma) identifié pour la première fois au Brésil, démontrant une augmentation de la transmissibilité dans le monde par rapport à B.1.1.7 (Alpha) et B.1.351 (Beta) de 10% (**Campbell et al., 2021**); infiltré en Algérie le 6 juillet 2021 en causant dans les environs de 13036 nouvelles infections et 6276 décès. Ce variant est notamment porteur de trois mutations dans la protéine Spike (E484K, N501Y et D614G) qui lui confèrent une transmissibilité et une virulence plus élevée qui ne peut être freinée par l'immunité naturelle apportée par une infection antérieure par la souche d'origine SARS-CoV-2 (**Larsen et al., 2020**). La mutation E484K présente déjà dans les variants Alpha, Beta et Delta de l'arbre du gène S est présente dans le variant Gamma par une séquence d'Uruguay. La mutation E484K augmente sensiblement la quantité d'anticorps sériques nécessaires pour empêcher l'infection des cellules (**Wise et al., 2021**), alors que la mutation D614G augmente la densité de la protéine S à la surface des particules virales et la liaison au récepteur ACE2 (**Zhang et al., 2020, Benvenuto et al., 2020**); cette dernière mutation est représentée dans l'arbre du gène S par un échantillon du Venezuela, en parallèle la mutation N501Y est représentée par une séquence nucléotidique d'Allemagne.

Cinquième vague : Variant Omicron

La cinquième vague est due au variant Omicron, signalé pour la première fois à l'OMS par l'Afrique du Sud; en Algérie a été signalée le 14 décembre 2021, le premier cas du variant une personne de nationalité étrangère qui est revenue sur le territoire algérien a subi un test antigénique à l'aéroport et le résultat s'est avéré positif. La vague du variant Omicron qui est plus contagieux mais moins dangereux que Delta, associé à un nombre record d'infections dans le monde et en Algérie (17119 est de 2503 décès enregistrés jusqu'en février 2022). Ce variant possède un nombre de mutations inhabituellement élevé, dont une trentaine dans la protéine spike, la clé d'entrée du virus dans l'organisme (**Nelson et al., 2021**), parmi ces mutations la mutation T478K représenté dans notre travail par une séquence d'Afrique du sud et la mutation N501Y représentées par deux séquence Bengladesh et Australie. La mutation T478K perturbe la fixation des anticorps monoclonaux sur la protéine S qui est due à la différence dans la longueur des chaînes latérales entre la Lysine (plus longue) et la Thréonine (le 478^{ème} acide aminé de la protéine S, une Tyrosine remplacé une lysine dans la séquence de l'Afrique du Sud) (**Nonaka et al., 2021**).

Résultats et Discussion

L'évolution de l'épidémie dans les différentes régions d'Algérie a été représentée sous forme graphique (Figure 47). L'Algérie a récemment enregistré une augmentation significative du nombre d'infections et de décès dus au virus corona émergeant, et selon les données de la courbe, le nombre d'infections exprimées de mars 2020 à mai 2022 a atteint plus de 100 000, et le nombre total de décès a atteint 10 000 (OMS, 2022).

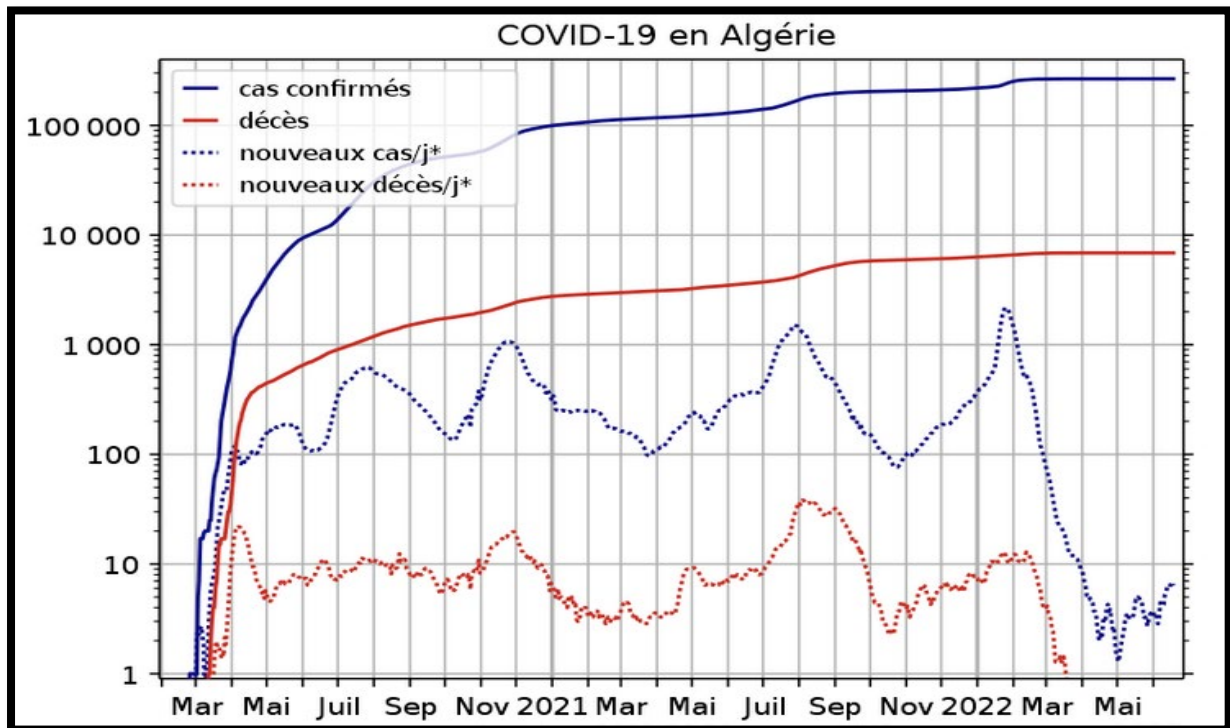


Figure 47 : une courbe montrant l'évolution du coronavirus en Algérie de 2020 à 2022 (OMS ,2022).

Depuis l'apparition sur le sol algérien du variant Omicron (cinquième vague), qui est plus contagieux mais moins dangereux que Delta, l'Algérie a lancé une nouvelle campagne anti-Covid. Cette campagne de vaccination en Algérie a débuté le 30 janvier 2021 au niveau de la Wilaya de Blida, avec pour premier vaccin utilisé le Sputnik V ou Astra Zeneca, Pfizer et Moderna, dans le cadre du programme covaxet d'après les données de l'OMS (2021), l'Algérie a vacciné 13 % de sa population générale contre le COVID-19, atteignant ainsi l'objectif mondial de l'OMS visant à élargir la vaccination et à aider à inverser le cours de la pandémie. Les principaux types de vaccins actuellement utilisés dans l'Algérie, en particulier les 2 vaccins à technologie ARNm Pfizer/BioNTech et Moderna sont le plus utilisé a récemment publié des résultats prometteurs indiquant que leurs vaccins sont efficaces à 94-95 % (Zeghibib et al., 2021).

CONCLUSION

Conclusion

Tous les virus y compris le SARS-CoV-2 mutent avec le temps, la plupart des mutations n'ont que peu ou pas d'incidence sur les propriétés du virus. Cependant certaines mutations sont plus inquiétantes et constituent un défi pour la prévention et le traitement du SARS-COV-2. La plupart des vaccins contre le SARS-COV-2 approuvés ou qui sont encore en cours d'essais cliniques sont conçus sur la base des gènes les plus variables S, ORF1ab, N, ORF3a, cependant la fréquence de mutation de ces gènes est très élevée car elle est en première ligne contre le système immunitaire ce qui l'oblige à s'adapter et à évoluer par des mutations lui conférant des capacités de neutralisation et d'évasion immunitaire qui rendent la tâche de confection des vaccins difficile.

A travers ce travail nous avons pu constater que le SARS COV-2 évolue par des mutations et des variants au niveau de son génome, ces mutations peuvent se produire sur la totalité du génome virale comme elles peuvent se produire dans certaines régions bien spécifiques du génome. La majorité des mutations détectées s'accumulent au niveau de ces variants: le variant anglais (Alpha B.1.1.7), le variant sud-africain (Béta B.1.351), le variant indien (Delta B.1.617.2), le variant brésilien (Gamma P.1) et le variant (Omicron B.1.1.529) et plus particulièrement dans les gènes les plus variables. Ces gènes du Coronavirus comptent des mutations réparties sur la totalité de son code génétique, parmi ces mutations : K417N/T, L452R, K478T, E484K, E484Q, N501Y, I2230T, P314L, T1001I, P4619L, G204R, R203K, P344S, Q57H, S26L, G251V ; sont celles qui ont eu le plus d'impact sur la virulence du virus, sur sa transmissibilité ainsi que sur sa capacité à échapper au système immunitaire, par conséquent les variants qui abritent ces mutations sont les plus dominants et les plus contagieux au monde tout particulièrement le variant Alpha, Beta, Delta, Gamma, Omicron et qui le sont actuellement.

L'avenir s'annonce prometteur compte tenu des progrès extraordinaires qui ont été réalisés dans la conception et le développement des vaccins Covid-19, alors que nous nous efforçons de rendre les vaccins accessibles aux personnes de tous horizons, y compris les moins développés, afin que nous puissions enfin empêcher la propagation d'Infection par le SARS-CoV-2.

Références bibliographiques

- **Amanat F., Krammer F.2020.** SARS-CoV-2 Vaccines: Status Report. *Immunity* 52, 583–589, P:7
- **Avise J., Arnold J., Ball M., Bermingham E., Lamb T., Neigel J., Reeb C., Saundres N. 1987.** “Intraspecific phylogeography: The mitochondrial DNA Bridge Between population genetics and systematics.” *Annual review of ecology and systematics* 18, 489-522, P: 34.
- **Beaulie M. 2007.** Caractérisation moléculaire des champignons ophiostomatoïdes associés à quatre espèces de scolytes de l'écorce colonisant l'épinette Blanche au Québec et phylogénie multigénique d'une nouvelle espèce de leptographium. Thèse doctorat, Université Laval, P:149.
- **Benevento D., Banu Demir A., Giovanetti M., Bianchi M., Angeletti S., Pascarella S., Cauda R., Antonio C.2020.** Evidence for mutations in SARS-CoV-2 Italian isolates potentially affecting virus transmission. *J Med Virol*, 92,2232–2237, P: 6
- **Benson D., Karsch I., Lipman D. J., Ostell J., Wheeler D. 2006.** GenBank. *Nucleic Acids Research*, (Issue suppl), D26–D31, P: 6.
- **Bérard S.2003.** Comparaison de séquences répétées en tandem et application à la génétique. Thèse de doctorat en Informatique. Thèse : Gascuel O. Univ. Montpellier II. 33-55, P: 242.
- **Berman H., Westbrook J., Feng Z., Gilliland G., Bhat T., Weissig H., Shindyalov I., Bourne E. 2000.** The Protein Data Bank. *Nucleic Acids Res*, 28(1), 235-242, P: 8.
- **Biao T., Nicola L., Qian L., Sanyi T., Yanni X., Jianhong W. 2020.** An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infectious Disease Modelling*, 248-255, P: 8.
- **Boccia A., Tufano R., Fetrucci V., Sepe L., Bianchi M., Pascarella S., Zollo M., Paoletta G. 2022.** SARS-CoV-2 Pandemic Tracing in Italy Highlights Lineages with Mutational Burden in Growing Subsets. Academic Editor: Chih-Jen Yang *Int. J. Mol. Sci*, 23(8), P: 26.

- **Boehm E., Kronig I., Neher R., Eckerle I., Vetter P., Kaiser L.2021.** Novel SARS-CoV-2 variants: the pandemics within the pandemic. *Clin Microbiol Infect*,1109-1117, P: 9.
- **Bontempi E.2020.** First data analysis about possible COVID-19 virus airborne diffusion. *Environmental Research* ,0013-9351,P: 8.
- **Chen H., Smith G., Li K., Wang J., Fan X., Rayner J. 2006.**Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. *The National Academy of Sciences of the USA*, 2845-2850, P:6.
- **Cheng C., Mombaerts L ., Wang M ., Hu T ., Sun C ., GuoY ., Xiuting L ., Hui Xu., Tongxin R., Yang X., Yaru X ., Hongling Z ., Honghan W ., Kezhi L ., Chuming C ., Yingxia L ., Zhichao L , Zhiguo C , Hai-Tao Z , Ioannis C., Paschaldis M ., Quanying Liu I ., Goncalves J ., Zhong Q ., Yan L. 2020.**Development and Validation of a Prognostic Risk Score System forCOVID-19 Inpatients: A Multi-Center Retrospective Study in China, *Engineering* 8,2095-8099, P: 6.
- **Cochrane G., Aldebert P., Althorpe N., Andersson M., Baker W., Baldwin A., Bates K., Bhattacharyya S., Bronwe P., Broek A., Castro M., Duggan K., Eberhardt R., Faruque N., Gamble J., Kanz C., Kulikova T., Lee C., Leinonen R., Lin Q., Lombard V., Lopez R., McHale M., William H., Mukherjee G., Nardone F., Pastor M. P., Sobhany S., Stoehr P., Tzouvara K., Vaughan R., Wu D., Zhu W., Apweiler R. 2006.** EMBL Nucleotide Sequence Database: developments in 2005. *Nucleic Acids Res*, (Issue suppl-1, 1) D10–D15, P: 6.
- **Comet J.1998.** Programmation Dynamique et Alignements de Séquences Biologiques. Thèse de doctorat en Informatique. Thèse : Henry J. Univ de Technologie de Compiegne, 5-38, P: 216.
- **Cheen Y., Lieu Q., Gueo D.2021.** Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol*1069 2020, 92, 418–423, P: 6.
- **Domingues M., Teixeira M ., Freitas M., Silva S., Tohidi., Zille J., Silva C., Felgueiras H. 2022.**Inhibition of Escherichia Virus MS2, Surrogate of SARS-CoV-2, via Essential Oils-Loaded Electrospun Fibrous Mats: Increasing the Multifunctionality of Antivirus Protection Masks. *Pharmaceutics*, 14, 303, P: 25.
- **Felsenstein J.1993.** PHYLIP (PHYLogeny Inference Package) version 3.6a2, Distributed by the author, Department of Genetics, University of Washington, Seattle, P: 50.

- **Felsenstein J.2004.** Inferring phylogenies. Sunderland (MA): Sinauer Associates: 664, P: 645.
- **Flores M., Anja S.2020.** The COVID-19 pandemic and its effects on teacher education. *Journal of Education for Teaching*, 453-456, P: 5.
- **Fung Y., Yuen K., Ye Z., Chan C., Jin A.2020.** Tug-of-war between severe acute respiratory syndrome coronavirus 2 and host antiviral defence: lessons from other pathogenic viruses. *Emerg Microbes Infect*, 9 (1), 558–70, P: 13.
- **Gabriel S., Maarten S., Dallas C., Yejin Ch., Noah A. 2019.**The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668-1678, P: 11.
- **Gao Q., Bao L., Mao H., Wang L., Xu K., Yang M., Li Y., Zhu L., Wang N., Lv Z.2020.** Development of an inactivated vaccine for SARS-CoV-2. *Science*369,77-81,P: 4.
- **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. *Microbiol Infect Dis*, 39(6)1011-1019,P: 9.
- **Greaney A., Starr T., Gilchuk P., Zost S., Binshtein E., Loes A. 2021.** Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe*, 29(1), 44-57, P: 13.
- **Guindon S. 2003.** Méthodes et algorithmes pour l'approche statistique en phylogénie, Thèse de Doctorat en Biologie. Univ. Montpellier II, 9-50, P: 155.
- **Haeckel E.1860.**Über neue, lebende Radiolarien des Mittelmeeres. *Monatsberichte der Kdniglichen*, 794-817, P: 23.
- **Hossam M., Walid F., Elkhatib M., Masudur Rahman., Hatem A. 2020.** Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks.*Pathogens*9, 186, P: 15.
- **Hoummadi J., Hafid S., Machraoui A. 2020.** Jusqu'ou` l'Afrique peut-elle limiter l'impact de la pandémie de COVID-19 ? revue d'Épidemiologie et de Santé Publique 68 (5) 302–305, P: 5.
- **Hozhabri H., Sparascio F., Sohrabi H., Mousavifar L., Roy R., Scribano D., De Luca A., Ambrosi C., Sarshar M.2020.**The Global Emergency of Novel Coronavirus (SARS-CoV-2): An Update of the Current Status and Forecasting. *Int. J. Environ. Public Health*, 17(16), 5648, P: 34.

- **Hang W., Lie X., Reen L., Zhao J., HueY. Cao B. 2020.** Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. 395, 497–506, P: 10.
- **Jaideep M., Senthil K., Purushothaman V.2020.** Increased mortality of COVID-19 infected diabetes patients: role of furin proteases. *International Journal of Obesity*, 44,2486-2488, P: 3.
- **Jubby M., Henry M., Chaparro S., Ludwig L.2021.** Mutation profile of SARS-CoV-2 genome in a sample from the first year of the pandemic in Colombia. 1567-1348, P: 8.
- **Kanyarat P., Darunee B., Piyawan Ch., Khajohn J., Wudtichai M., Angkana H., Thitiman K., Narong M., b Rotjana K., Sopon I., Anthony R., Aaron R. 2021.** Coding-Complete Genome Sequences of Alpha and Delta SARS- CoV-2 Variants from Kamphaeng Phet Province Thailand. *Microbiol Resour Announc* 10:e00877-21, P: 9.
- **Kemp S., Collier D., Datir R., Ferreira I., Gayed S., Jahun A., Hosmillo M., Rees-Spear C., MlcochovaInes P., Lumb U., Roberts D., Chandra A., Temperton N. 2020.** Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. Preprint, P: 52.
- **Ketfi A., Chabati O., Chemali S., Mahjoub M., Gharnaout M., Touahri R., Djenouhat K., Selatni F., Ben Saad H.2020.** Profil Clinique, biologique et radiologique des patients Algériens hospitalisés pour COVID 19: données préliminaires. *Pan Arica Médical Journal*, 35(2), 1937-8688, P:28.
- **Konings F., Campbell F., Archer B., Laurenson H., Jinnai Y., Vandemaele K., Morgan O.2021.** Increased transmissibility and global spread of SARS-CoV-2 variants. *Euro Surveill*, 26(24), P: 6.
- **Koyama T., Platt D., Parida L.2020.** Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ* 98(7),495-504, P: 9.
- **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):1547-1549,P: 3.
- **Kannan S., Spratt A., Cohen A., Naqvi S., Chand H., Quinn T. 2021.** Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *J Autoimmune*, 124-102715, P: 5.
- **Laamarti M., Chemaou M., Laamarti R., Allam L., Ouadghiri M., Smyej I., Rahoui J., Benrahma H., Diawara I., Alouane T., Essabbar A., Siah S., Karra M., El Hafidi N., El JaoudiR., Sbabou L., Nejari C., Amzazi S., Mentag R., Belyamani L., Ibrahimi A. 2020.** Genomesequences of six SARS-CoV-2 strains isolated in Morocco

obtained using Oxford Nanopore MinION technology. *Microbiol Resour Announc* 9(32), e00767-20, P: 8.

- **Larsen C., Paludan S. 2020.** Corona's new coat: SARS-CoV-2 in Danish minks and implications for travel medicine. *Travel Med Infect Dis* 38, P: 3.
- **Li K. 2003.** ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics Institute*, 1585-1586, P: 2.
- **Luan B., Wang H., Huynh T. 2021.** Molecular Mechanism of the N501Y Mutation for Enhanced Binding between SARS-CoV-2's Spike Protein and Human ACE2 Receptor. *FEBS Lett*, 595(10), 1454-1461, P: 8.
- **Lucy van D., Richard D., Cedric T., Liam P. 2020.** No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Environmental Research* 186, 109639, P: 8.
- **Lutteropp S., Ucas L., Lexandros S., Ortiz M., Ntonis R. 2019.** Quartet-Based Computations of Internode Certainty Provide Robust Measures of Phylogenetic Incongruence. *Syst Biol*, 69 (2), 308-324, P: 14.
- **Ma, C., Su S., Wang J., Wei L., Du L., Jiang S. 2020.** From SARS-CoV to SARS-CoV-2: Safety and broad-spectrum are important for Coronavirus vaccine development. *Microbes Infect.* 22(6), 245-253, P: 9.
- **Mayr E. 1965.** Classification and phylogeny. *Amer. Zool*, 5, 165-174, P: 9.
- **Meriadeg Le gouil., Éric Maurice Leroy., Jeanne B. 2020.** Transmission du Covid-19 aux animaux de compagnie : un risque à ne pas négliger. In: *Bulletin de l'Académie Vétérinaire de France* tome 173 n°1, 71-74, P: 5.
- **Mohd I., Zainul A., Hassan M., Irfan A., Abd-ElAzim F. 2020.** Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight. *PLoS ONE* 15 (9), e0238344, P: 18.
- **Moreira R., Chwastyk M., Baker J., Guzman H., Poma A. 2020.** Quantitative determination of mechanical stability in the novel coronavirus spike protein. *Nanoscale*, 12, 16409-16413, P: 5.
- **Muchtaridi M., Fauzi M., Khairul Ikram N., Mohd G., Wahab H. 2020.** Natural Flavonoids as Potential Angiotensin-Converting Enzyme 2 Inhibitors for Anti-SARS-CoV-2. *Molecules*, 25, (17), 3980, P: 20.

- **Mulder N., Schwartz R., Brazas M., Brooksbank C., Gaeta B., Morgan S. 2018.** The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLoS Comput Biol*, 14(2), e1005772, P: 14.
- **Muñoz-Fontela C., Dowling W., Funnell S., Gsell P., Riveros-Balta A., Albrecht R., Andersen H., Baric R., Carroll M., Cavaleri M. 2020.** Animal models for COVID-19. *Nature*, 586, 509-515, P: 6.
- **Nicolás Gómez-Banoy., Jennifer D., Luis G. Gomez E., Lunkun M., Ankit G., Sergio A., Evan T., Vasuretha Ch., Yaron B., Katherine H., Priya B., Phoebe Pi., Alfonso Skyler., Lucia C., Sean H., Redmond D., Alpana P., Parag G., Kristy A., Benjamin R., Laura C., Robert E., Edward J., Monika M., James C. 2021.** Clinical and Translational Report Hyperglycemia in acute COVID-19 is characterized by insulin resistance and adipose tissue infectivity by SARS-CoV-2. *Cell Metabolism* 33, 2174-2188, P: 20.
- **Nonaka C., Franco M., Gräf T., Mendes A., Aguiar R., Giovanetti M. 2021.** Genomic Evidence of a Sars-Cov-2 Reinfection Case with E484K Spike Mutation in Brazil. *Preprints*, P: 6.
- **Nelson G., Buzko O., Spilman P., Niazi K., Rabizadeh S., Soon-Shiong P. 2021 .** Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone. Potentially resulting in an escape mutant, P: 13.
- **Okubo K., Sugawara H., Gojobori T., Tateno Y. 2006.** DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Res*, 34(Database issue), D6–9, P: 4.
- **Paula D., Halpert G., Shoenfeld Y. 2021.** The SARS-CoV-2 as an instrumental trigger of autoimmunity. *Autoimmunity Reviews* 20, 102792, P: 9.
- **Pirovano W., Heringa J. 2008.** Multiple Sequence Alignment. In: Keith JM, editor. *Bioinformatics: Data, Sequence Analysis and Evolution*. Totowa, NJ: Humana Press, 143-161, P: 18.
- **Pruitt K., Tatusova T., Maglott D. 2005.** NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, 33 (suppl_1), D501-D504, P: 4.

- **Ray M, Sable M, Sarkar S, Hallur V. 2021.** Essential interpretations of bioinformatics in COVID-19 pandemic. *Meta Gene* 27, 100844, P: 9.
- **Rischkowsky B., Pilling D. 2008.** L'Etat des Ressources Zoogénétiques pour l'Alimentation et l'Agriculture dans le Monde. *Presses de Sciences Po*,(pp393-410),P:440.
- **Ron A., M Fouchier., Vincent M., Anders W., Theo M., Sander H., Derek S., Guus F., Börjn O., Osterhaus A. 2005.** Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black- Headed Gulls. *J. Virol.* 79(5), 2814-2822, P: 8.
- **Schmidt A. 2003.** Phylogenetic Trees from Large Datasets Inaugural. Thèse de Doctorat en Mathématique. Von Haeseler A. Univ. Heinrich Heine Düsseldorf vorgelegt von Heiko, P: 123.
- **Servant F., Bru C., Carrere S., Courcelle E., Gouzy J., Peyruc D., Kahn D. 2002.** ProDom: automated clustering of homologous domains. *Brief Bioinform*; 246-251, P: 6.
- **Segondy M. 2020.** Les Coronavirus humains. *Revue Francophone Laboratoires* ;32-39, P: 8.
- **Sharma O., Sultan A., Ding H. 2020.** A Review of the Progress and Challenges of Developing a Vaccine for COVID-19, *Front. Immunol* 11, 2413, P: 17.
- **Shi J., Wen Z., Zhong G., Yang H., Wang C., Huang B., Liu R., He X., Shuai L., Sun Z. 2020.** Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. *Science* 368, 1016-1020, P: 4.
- **Shubham S., Suhas T., Mhaske S., Modak G., Shamburaje S., Akhilesh Ch., Vidya A. 2022.** Emergence of two distinct variants of SARS-CoV-2 and an explosive second wave of COVID-19: the experience of a tertiary care hospital in India, *Archives of Virology*, 167 (2), 393-403, P: 11.
- **Sunnucks P. 2001.** Efficient genetic markers for population biology. *Tree*, 15, 199-203, P: 4.
- **Swofford D., Olsen G., Waddell P., Hillis D. 1996.** Phylogenetic inference. In D. M Hillis, C. Moritz, and B. Mable (eds.) *Molecular Systematics* (2nded.), Sinauer Associates, Sunderland, Massachusetts, 407-514, P: 7.
- **Selene Z., Taboada B, Esteban J., Medina M., Iša P., Sanchez-Flores A., Boukadida C., Herrera-Estrella A., Selem N. 2022.** The Alpha Variant (B.1.1.7) of SARS-CoV-2 Failed to Become Dominant in Mexico. *Microbiology Spectrum* , P : 15.

- **Taboada B., Zárata S., Iša P., Boukadida C., Joel A., José E., Comas-García A. 2020.** Genetic Analysis of SARS-CoV-2 Variants in Mexico during the First Year of the COVID-19 Pandemic. Academic Editors: Eurico Arruda, Helena Lage Ferreira, José Luiz Proença Módena and William Marciel de Souza *Viruses*, 13(11), P: 20.
- **Tobias J., Lunn S., Miller T., Thorough D., Kevin B. 2021.** COVID-19 Outbreak Associated with a SARS-CoV-2 R.1 Lineage Variant in a Skilled Nursing Facility after Vaccination Program. *MMWR Morb Mortal Wkly Rep*, 70(17), 639-643, P: 11.
- **Tourasse N. 1992.** Développement d'une distance évolutive entre séquences prenant en compte la variabilité du taux de substitution entre sites et application à la reconstruction de phylogénies moléculaires anciennes. Thèse Doctorat en Génétique et Biologie des Populations, Univ Claude Bernard - Lyon 1, P: 122.
- **Wise J. 2021.** Covid-19: The E484K mutation and the risks it poses. *BMJ* 372,n359,P: 2.
- **World Health Organization 2020.** Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). WHO: Geneva, Switzerland: 40.
- **Wu C., Apweiler R., Bairoch A., Natale D., Barker W., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M., Mazumder R., O'Donovan C., Redaschi N., Suzek B. 2006.** The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*, 34(Database issue), D187–91,P: 6.
- **Wu I., Jianyun L., Yanhui L., Zhoubin Z. 2020.** Positive effects of COVID-19 control measures on influenza. Elsevier Ltd on behalf of International Society for Infectious Diseases, P: 2.
- **Xu J., Zhao S., Teng T., Abdalla A., 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.P:17
- **Xia X., Xie Z. 2001.** DAMBE: Software package for Data Analysis in Molecular Biology and Evolution. *Computer Notes* 373, P: 3
- **Xia X. 2013.** Department of Bio. DAMBE5: A Comprehensive Software Package for Data Analyses in Molecular Biology and Evolution. Department of Biology and Center for Advanced Research in Environmental Genomics. University of Ottawa, Ottawa Ontario, Canada, P: 9
- **Zhang Wenjuan., Brian D., Stephanie S. 2021.** Emergence of a Novel SARS-CoV-2 Variant in Southern California,325(13),1324-1326, P: 3.

- **Zhang, H., Penninger, J.M., Li, Y., Zhong, N., Slutsky, A. 2020.** Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: Molecular mechanisms and potential therapeutic target. *Intensive Care Med*, 46, 586-590.P: 4.
- **Zeghib S., Balázs A., Brigitta Z., Kemenesi G., Herczeg R., Derrar F ., Jakab F. 2021.** The Algerian Chapter of SARS-CoV-2 Pandemic: An Evolutionary, Genetic, and Epidemiological Prospect. *Viruses* 13(8), 1525, P: 20.
- **Zhou Y., Lu L., Wei Z., Chun C., Yao H., Jiaying S., Zhijie L., Weiqi X., Qingqing W., Dongsheng J., Pingqing F., Zifa W., Douglas R. 2020.**A chemical cocktail during the COVID-19 outbreak in Beijing, China: Insights from six-year aerosol particle composition measurements during the Chinese New Year holiday. *Science of the Total Environment* 742, 140739, P: 11.

Biblio-web

- **Blast:** <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- **FaBox:** <https://users-birc.au.dk/palle/php/fabox/>
- **MEGA:** <https://www.megasoftware.net/>
- **NCBI :** <https://www.ncbi.nlm.nih.gov/genbank>
- **OMS:** <https://www.who.int/fr/>