

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

**Inventaire et traitement par les outils bioinformatiques des gènes identifiés
chez le Coronavirus, exprimé cliniquement par un syndrome de détresse
respiratoire aiguë (SDRA).**

Présenté par :

BENZAYED Yasmine

BELMIHOUB Maroua

Devant le jury :

Mme ROUAKI F.

M.C.A USDB 1

Présidente

Mme ABDULHUSSEIN A.

M.C.A USDB 1

Examinatrice

Mme DEROUICHE L.

M.C.A ESSAIA

Promotrice

Mer MOHAMED SAID R.

M.C.A USDB 1

Co-promoteur

Promotion: 2020/2021

REMERCIEMENTS

En préambule à ce mémoire, nous remercions **ALLAH** qui nous a aidé et donné la patience, le courage et la volonté durant cette longue année d'étude.

Nous sommes très heureuses d'exprimer ici tous nos remerciements et gratitudes à notre promotrice

Mme DEROUICHE L.

Maitre de conférences à L'École Supérieure des Sciences de l'Aliment et Industries Agroalimentaires–Alger, pour sa disponibilité tout au long de la réalisation de ce mémoire, ainsi que pour son inspiration, remarques et conseils, aide et suivi.

Nous remercions très chaleureusement aussi, *Mr MOHAMED SAID R.* notre co-promoteur et chef d'option génétique, pour sa confiance et ses encouragements.

Nos remerciements iront également vers tous ceux qui ont accepté avec bienveillance de participer au jury de ce mémoire :

La présidente *Mme ROUAKI F.* d'avoir bien voulu nous faire l'honneur de présider ce jury.

L'examinatrice *Mme ABDULHUSSEIN A.* pour avoir accepté d'évaluer ce travail et pour toutes leurs remarques et critiques.

Nous remercions, *Mr BEGHDED Amine* informaticien à l'École nationale préparatoire aux études d'ingénieur.

Merci à tous



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 devenu ce travail est de lui, grâce à lui. Louanges et remerciements
pour votre clémence et miséricorde*

À Ma Chère Mère Salima

*Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma
considération pour les sacrifices que vous avez consenti pour mon instruction et mon
bien être.*

*Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon
enfance et j'espère que votre bénédiction m'accompagnera toujours. Que ce modeste
travail soit l'exaucement de vos vœux tant formulés, le fruit de vos innombrables
sacrifices. Puisse Dieu, le très haut, vous accorde santé, bonheur et longue vie*

À La Mémoire de Mon Père Ali

*Ce travail est dédié à mon père, décédé trop tôt. J'espère que, du monde qui est sien
maintenant, il apprécie ce humble geste comme preuve de reconnaissance de la part
d'une fille qui a toujours priée pour le salut de son âme. Puisse Dieu le tout puissant,
l'avoir en sa sainte miséricorde*

À mes chers frères Amine et Ali Riadh

*A tous les moments d'enfance passés avec vous mes frères, en gage de ma profonde
estime pour l'aide que vous m'apporterez. vous me soutenez, réconfortez et
encouragez. Puisse nos liens fraternels se consolider et se renforcer encore plus.*

À Ma Chère Famille

*Mon grand père Abdellah et ma grand mère Yamouna. Ma chère tante unique
Fatmazohra et son mari Abd Elhak, mes chères oncles Adlene et Djamel et leurs
femmes, mes meilleurs cousins, Aimen, Rihab, Djihan, Ferial, Hadjer, j'espère que
mon mémoire sera pour vous une source de fierté, je vous souhaite une vie pleine de
bonheur et de succès et que Dieu, le tout puissant, vous protège et vous garde.*

À Mes chers Amis

*Ma chère binome Maroua, Safa, Sabrina, Sarah. A travers mon travail, je vous
transmets mes meilleurs sentiments
d'amour.*

Yasmine

Merci à tous



Dédicaces

Toutes les lettres ne sauraient trouver les mots qu'il faut... Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance...

Aussi, c'est tout simplement que

Je remercie Dieu

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

À Mes Très Chers Parents

Tous les mots du monde ne sauraient exprimer l'immense amour que je vous porte, ni la profonde gratitude que je vous témoigne pour tous vos efforts, et vos sacrifices que vous n'avez jamais cessé de consentir pour mon instruction et mon bien-être. Je vous dédie ce travail et je vous remercie pour tout ce que vous avez fait pour moi. Puisse Dieu vous préserver et vous procurer santé, bonheur et longue vie et faire en sorte que jamais je ne vous déçoive.

À Ma Chère Famille

Mon frère, mes chères sœurs, ma belle-sœur, mes nièces et mes neveux

Nous avons partagé les meilleurs et les plus agréables moments, votre amour, énergie, joie et sourire ont été pour moi le meilleur encouragement que je puisse avoir. J'espère que mon mémoire sera pour vous une source de fierté, je vous souhaite une vie pleine de bonheur et de succès et que Dieu, le tout puissant, vous protège et vous garde.

À Toutes Mes Amies Et Tous Mes Amis

*Ma chère sœur **Hiba**, mon amie palestinienne **Aesha** et mon très cher binôme **Yasmine**.*

En témoignage de l'amitié qui nous unie et des souvenirs de tous les moments que nous avons passé ensemble.

A travers mon travail, je vous transmets mes meilleurs sentiments d'amour.

Que Dieu vous donne longue vie pour le maintien de l'union de notre lien sacré.

Je vous en serais toujours reconnaissante et soyez assurés de mon estime et mon profond respect.

Merci à tous

Maroua



Tableaux et figures

Liste des tableaux

N° de tableau	Titre	Page
I	Classification de Covid 19 (Zi-Wei et Shuofeng, 2020 ; Patrick et al., 2020 ; Kapikian, 2020).	1
II	Description morphologique et génomique de Covid 19 (Sukhadeo et al., 2020 ; Hong et al., 2020 ; Korsia-Meffre, 2021 ; Laib, 2021).	2
III	Statistiques de Covid 19 dans le monde et dans quelques pays (Moyou, 2021).	6
IV	SARS Covid 2 en Algerie (Dib, 2020 ; Kourta, 2021 ; Fourar, 2020 ; Laib, 2021).	7
V	Les différents pays faisant des analyses des gènes de SARS Cov 2.	37
VI	Résumé des résultats des arbres phylogénétiques obtenus par MEGA des gènes de SARS Cov2.	41
VII	Les résultats de DnaSP des gènes de SARS Cov2.	43
VIII	Les variations globales AT et CG Skew des gènes S, N, ORF3a et ORF1ab de SARS Cov 2.	47

Liste des figures

N° de figure	Titre	Page
01	Micrographie électronique à transmission couleur (TEM) d'une particule de Coronavirus SRAS Cov 2 (Hopkins, 2021).	1
02	Présentation de Covid19 (Hong et al., 2020).	3
03	Cycle virale de SARS Covid 19 (Purva et al., 2020).	4
04	Les approches diagnostiques et thérapeutiques de Covid 19 (Chanthadavong, 2020).	6
05	Arbre phylogénétique non enraciné (Benhamdi, 2020).	11
06	Arbre phylogénétique enraciné (Benhamdi, 2020).	11
07	Approches diagnostiques et thérapeutiques de Covid19 (Mahalaxmi et al.,	20

Tableaux et figures

	2020).	
08	Equipement de laboratoire de biologie moléculaire utilisé pour l'analyse de RT-PCR de SARS Cov 2 (Aoues, 2020).	21
09	Equipement de protection individuelle dans laboratoire d'analyse de SARS Cov 2 (Aoues, 2020).	21
10	Les étapes initiales pour récupérer les séquences de la GenBank.	23
11	Analyse de genre « SARS Cov 2 » par la GenBank.	24
12	Tableau informatif de SARS Cov 2 obtenue par la GenBank.	24
13	Exemple des résultats nucléotidiques de SARS Cov 2 obtenue par la GenBank.	25
14	Exemple d'une séquence nucléotidique sous format FASTA dans la GenBank.	25
15	La page principale du programme FaBox.	26
16	Conversion des séquences en forme FASTA par l'utilisation de FaBox.	27
17	Barre d'outils de MEGA 11.	28
18	L'ajout des séquences nucléotidiques sur logiciel MEGA 11.	28
19	L'alignement multiple des séquences nucléotidiques par le programme ClustalW.	29
20	La méthode UPGMA utilisée pour la construction de l'arbre phylogénétique.	29
21	Les paramètres utilisés pour la reconstruction phylogénétique.	21
22	Barre d'outils de DnaSP v6.	31
23	Interface utilisateur de DnaSP v6.	31
24	Barre d'outils du programme DAMBE v7.3.2.	32
25	L'étape d'obtention des résultats statistiques et graphiques de CG/AT.	33
26	La page d'accueil de Network v10.2.0.0.	34
27	Ouverture du fichier de données de rdf pour le calcul de MJ NetWork.	34
28	Enregistrement d'un fichier de calcul MJ sous forme out .	35
29	Les dernières étapes de reconstruction du réseau haplotypique.	35
30	Page d'accueil en ligne de programme FAS to RDF Conversion.	36
31	Un exemple de forme FASTA obtenue par FaBox.	40
32	Résultat du programme Adn Fasta format-MEGA v1.0.	41

Tableaux et figures

33	Arbre phylogénétique de gène ORF1ab de SARS Cov 2 construit à partir de 37 séquences, en utilisant la méthode UPGMA par logiciel MEGA.	42
34	Arbre phylogénétique de gène ORF 6 de SARS Cov 2 construit à partir de séquences, en utilisant la méthode UPGMA par logiciel MEGA.	43
35	Nombre des haplotypes de gène ORF1ab calculé par le logiciel DnaSP.	44
36	Nombre des haplotypes de gène E calculé par le logiciel DnaSP (haplotype unique).	45
37	Analyse des bases azotiques des séquences de gène E par DnBA.	46
38	Analyse des bases azotiques des séquences de gène S par DnBA.	46
39	Résultats DAMBE, les séquences de paires identiques de gène ORF1ab de SARS Cov 2.	48
40	Les analyses statistiques d'une séquence de gène d'ORF1ab par logiciel DAMBE.	48
41	La variation de GC et AT skews de gène ORF1ab (séquence de Vietnam) obtenu par logiciel DAMBE.	48
42	Réseau phylogénétique des séquences de Gene ORF1ab de SARS Cov2, reconstruit par logiciel NetWork à partir de 37 séquence, représentés par des cercles colorés, chaque couleur représente un pays.	49
43	Réseau phylogénétique de séquences de Gene ORF1ab de SARS Cov2, reconstruit par NetWork à partir de 37 séquences, représentées par des cercles colorés, chaque couleur représente un pays.	50
44	Arbre phylogénétique du gène S de SARS Cov 2 construit à partir de 37 séquences d'ADN, en utilisant la méthode UPGMA par logiciel MEGA.	Annexe
45	Réseau phylogénétique de séquences de Gene S de SARS Cov 2, reconstruit par le logiciel NetWork à partir de 37 séquences, représentées par des cercles colorés, chaque couleur représente un pays.	Annexe
46	Variations du GC skew de gène S réalisées par le logiciel DAMBE.	Annexe
47	Variations du AT skew de gène S réalisées par le logiciel DAMBE.	Annexe
48	Arbre phylogénétique de gène ORF3a de SARS Cov 2 construit à partir de 41 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe
49	Réseau phylogénétique de séquences de gène ORF3a de SARS Cov2, reconstruit par le logiciel NetWork à partir de 41 séquences, représentées par des cercles colorés, chaque couleur représente un pays.	Annexe

Tableaux et figures

50	Analyse des bases azotiques des séquences de gène ORF3a par DnBA.	Annexe
51	Variations du GC skew de gène ORF3a réalisées par le logiciel DAMBE.	Annexe
52	Variations de AT skew de gène ORF3a réalisées par logiciel DAMBE.	Annexe
53	Arbre phylogénétique de gène N de SARS Cov 2 construit à partir de 43 séquences, en utilisant la méthode UPGMA par logiciel MEGA.	Annexe
54	Réseau phylogénétique de séquences de Gene N de SARS Cov2, reconstruit par le logiciel NetWork à partir de 43 séquences, représentées par des cercles colorés, chaque couleur représente un pays.	Annexe
55	Analyse des bases azotiques des séquences de gène N par DnBA.	Annexe
56	Variations du GC skew de gène N réalisées par le logiciel DAMBE.	Annexe
57	Variations du AT skew de gène N réalisées par le logiciel DAMBE.	Annexe
58	Analyse des bases azotiques des séquences de gène ORF1ab par DnBA.	Annexe
59	Variations du GC skew de gène ORF1ab réalisées par le logiciel DAMBE.	Annexe
60	Variations du AT skew de gène ORF1ab réalisées par le logiciel DAMBE.	Annexe
61	Arbre phylogénétique de gène ORF10 de SARS Cov 2 construit à partir de 39 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe
62	Réseau phylogénétique de séquences de Gene ORF10 de SARS Cov 2, reconstruit par le logiciel NetWork à partir de 39 séquences, représentées par des cercles colorés, chaque couleur représente un pays.	Annexe
63	Analyse des bases azotiques des séquences de gène ORF10 par DnBA.	Annexe
64	Arbre phylogénétique de gène ORF8 de SARS Cov 2 construit à partir de 37 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe
65	Analyse des bases azotiques des séquences de gène ORF8 par DnBA.	Annexe
66	Arbre phylogénétique de gène ORF7b de SARS Cov 2 construit à partir de 26 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe Annexe
67	Analyse des bases azotiques des séquences de gène ORF7b par DnBA.	Annexe
68	Analyse des bases azotiques des 2 séquences de gène ORFa1 par DnBA.	Annexe
69	Arbre phylogénétique de gène ORF7a de SARS Cov 2 construit à partir de 39 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe
70	Analyse des bases azotiques des séquences de gène ORF7a par DnBA.	Annexe
71	Arbre phylogénétique de gène E de SARS Cov 2 construit à partir de 43 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe

Tableaux et figures

72	Arbre phylogénétique de gène M de SARS Cov 2 construit à partir de 40 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe
73	Analyse des bases azotiques des séquences de gène M par DnBA.	Annexe
74	Arbre phylogénétique de gène ORF6 de SARS Cov 2 construit à partir de 41 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.	Annexe
75	Analyse des bases azotiques des séquences de gène ORF 6 par DnBA.	Annexe

Abréviations

- **ACE2** : Enzyme de conversion de l'angiotensine 2.
- **ADN**: Acide DésoxyriboNuc léique.
- **ADNmt** : ADN mitochondrial.
- **ARA 2** : Antagonistes des Récepteurs de l'Angiotensine 2.
- **ARN**: Acide Ribo Nuc léique.
- **DAMBE**: Data Analysis in Molecular Biology and Evolution.
- **DnaSP**: DNA Séquence Polymorphism.
- **DnBA** : DNA Bases Analysis .
- **EBI** : European Bioinformatics Institute.
- **ECMO** : Oxygénation Extracorporelle de la Membrane.
- **EHS** : Etablissement Hospitalier Spécialisé.
- **ELISA** : Acronyme de Enzyme Linked ImmunoSorbent Assay .
- **EPI** : Equipements de Protection Individuelle.
- **FaBox** : FASTA sequence toolBOX .
- **HTA** : Hypertension Artérielle.
- **IEC** : Inhibiteur de l'Enzyme de Conversion de l'angiotensine.
- **LEAPdb** : Late Embryogenesis Abundant Proteins database.
- **LFA** : Lateral Flow Assay.
- **MEGA**: Molecular Evolutionary Genetics Analysis using Maximum Likelihood.
- **MERS Cov** : Coronavirus du Syndrome Respiratoire du Moyen-Orient.
- **ML** : Maximum Likelihood.
- **NCBI** : National Center for Biotechnology Information.
- **NJ** : Neighbor Joining.
- **OMS** : Organisation Mondiale de la Santé.
- **ORF** : Open Reading Frame.
- **PDB** : Protein Data Bank.
- **RDRP** : RNA-Dependent RNA Polymerase.
- **RE** : Réticulum Endoplasmique.
- **RT-PCR** : Reverse Transcription, Réaction de Polymérisation en Chaîne.

- **SARS Cov 2** : Coronavirus 2 du Syndrome Respiratoire Aigu Sévère.
- **SDRA** : Syndrome de Détresse Respiratoire Aigu.
- **sHSPdb** : Small Heat Shock Proteins database .
- **TEM** : Micrographie Electronique à Transmission couleur.
- **UPGMA**: Unweighted Pair Group Method with Arithmetic mean.

Glossaire

- **Bootstrap** : Méthode statistique utilisée pour évaluer la robustesse d'un arbre phylogénétique. Elle part de l'hypothèse que les sites évoluent de manière indépendante.
- **FASTA** : Algorithme proposé par Pearson basé sur l'identification rapide de fragments strictement identiques entre la séquence cible et les séquences de la banque (**Comet, 1998**).
- **Homologie** : Similitude héritée d'un ancêtre commun, contrairement à l'homoplasie (**Lankaster, 1870**).
- **Maximum de parcimonie** : Approche statistique non-paramétrique très utilisée, notamment pour l'inférence phylogénétique. Sous l'hypothèse du maximum de parcimonie, l'arbre phylogénétique inféré est celui qui requiert le plus petit nombre de changements évolutifs.
- **Maximum de vraisemblance** : Approche statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné. Elle est très employée dans le domaine de la bioinformatique.
- **Transition** : Mutation au cours de laquelle une base purique est remplacée par une autre base purique ou par une base pyrimidique.
- **Transversion** : Mutation au cours de laquelle une base purique est remplacée par une base pyrimidique ou vice-versa.
- **UPGMA** : Méthode utilisée pour construire des arbres phylogénétiques si les séquences ne sont pas trop divergentes (**Golding et al., 2003**).

RÉSUMÉ

Le coronavirus exprimé cliniquement par un syndrome de détresse respiratoire aiguë (SDRA), est un virus qui cause une pandémie mondiale, affectant des millions de personnes et mettant fin à la vie de plus de trois million de personnes à travers le monde. L'objectif de notre travail est d'inventorier les gènes étudiés chez ce virus ainsi que les différentes mutations détectés jusqu'à présent dans plusieurs pays, pour cela nous avons utilisé des outils de bioinformatique afin de réaliser des analyses phylogénétiques pour la construction des arbres et des réseaux phylogénétiques et faire des traitements statistiques des données moléculaires disponibles sur la GenBank concernant le COVID 19. Les différents logiciels utilisés sont : MEGA 11 pour la construction des arbres phylogénétiques, DnaSP6 pour l'identification des haplotypes, Network 10 pour avoir les réseaux phylogénétiques, DnBA un nouveau programme que nous avons créé pour réaliser des analyses statistiques, puis DAMBE pour obtenir les résultats moléculaire, statistiques et graphiques.

L'étude des séquences d'ADN obtenues de la GenBank, nous a permis de classer 12 gènes différents de SARS Cov 2 déterminé à travers les laboratoires du monde. L'utilisation des différents logiciels en se basant sur la détermination des mutations ainsi que la diversité haplotypique nous a permis de classer ces 12 gènes en trois catégories ; les gènes les plus variables : ORF1ab, ORF3a, N, S ; les gènes les moins variables : ORF8, ORF10, ORF1a, ORF7a, ORF7b et les gènes non variés : ORF6, E, M. Les gènes de la 1^{ère} et la 2^{ème} catégorie exposent des pourcentages A, T, C, et G différents avec un taux de GC bas par rapport à la 3^{ème} catégorie donc ils sont moins conservés , pour cela ils subissent des mutations. En classant les séquences d'ADN des 12 gènes étudiés selon les pays d'origine nous avons trouvé que la Chine est l'origine de toutes les séquences. Les résultats d'analyses phylogénétiques, statistiques et graphiques des différents logiciels utilisés montre qu'il y a une complémentarité entre ces différents outils de bioinformatique.

Mots clés : SARS Cov 2, moléculaire, mutations, phylogénétique, logiciels.

ملخص

يُعد فيروس كورونا الذي تم التعبير عنه سريريًا من خلال متلازمة الضائقة التنفسية الحادة (ARDS) ، فيروسًا يسبب وباءً عالميًا يؤثر على ملايين الأشخاص وينهي حياة أكثر من ثلاثة ملايين شخص حول العالم. الهدف من عملنا هو جرد الجينات التي تمت دراستها في هذا الفيروس بالإضافة إلى الطفرات المختلفة التي تم اكتشافها حتى الآن في العديد من البلدان ، ولهذا استخدمنا أدوات المعلوماتية الحيوية لإجراء تحليلات علم الوراثة لأشجار البناء وشبكات النشوء والتطور وإجراء معالجة إحصائية تتوفر على البيانات الجزيئية في GenBank بخصوص COVID 19.

تم تطوير تحليلات النشوء والتطور التي تم الحصول عليها في هذا العمل باستخدام برامج مختلفة مثل MEGA 11 لبناء أشجار النشوء والتطور باستخدام طرق مختلفة مثل UPGMA ، ومعلومات Kimura-2 ، و DnaSP6 لتحديد الأنماط الفردانية ، و NetWork 10 للحصول على النشوء والتطور بالإضافة إلى DnBA وهو برنامج جديد أنشأناه لإجراء التحليلات الإحصائية ، ثم DAMBE للحصول على نتائج جزيئية وإحصائية ورسوميات.

سمحت لنا دراسة تسلسل الحمض النووي التي تم الحصول عليها من GenBank بتصنيف 12 جينًا (ORF1ab ، ORF3a ، ORF8 ، S ، N ، ORF10 ، ORF1a ، ORF7a ، ORF7b ، ORF6 ، E ، M) مختلفة عن SARS Cov 2 في جميع أنحاء العالم. سمح لنا استخدام البرامج المختلفة على تحديد الطفرات بالإضافة إلى تصنيف هذه الجينات الـ 12 إلى ثلاث فئات: الجينات الأكثر تغيرًا ORF1ab : ORF3a ، N ، S ؛ الجينات الأقل تغيرًا ORF8 ، ORF10 ، ORF1a ، ORF7a ، ORF7b ، الجينات غير المتغيرة ORF6 ، E ، M ، تظهر نتائج التحليلات التطورية والإحصائية والرسوم البيانية لمختلف فئات البرامج المستخدمة أن هناك تكامل بين أدوات المعلوماتية الحيوية المختلفة

المصطلحات: SARS Cov 2 , الجزيئية , الطفرات , شجرة النشوء , البرمجيات .

Summary

The coronavirus expressed clinically as acute respiratory distress syndrome (ARDS), is a virus that causes a global pandemic, affecting millions of people and ending the lives of more than three million people worldwide. The objective of our work is to inventoried the genes studied in this virus and the different mutations detected so far in several countries, for this we used bioinformatics tools to perform phylogenetic analyses for the construction of phylogenetic trees and networks and to perform statistical treatments of the molecular data available in the GenBank concerning COVID 19.

The phylogenetic analyses obtained in this work were elaborated with the help of different software such as, MEGA 11 for the construction of phylogenetic trees with the use of different methods like UPGMA, Kimura- 2 parameters, DnaSP6 for the identification of haplotypes, Network 10 to have the phylogenetic networks, DnBA a new program that we have created to perform statistical analyses, and then DAMBE to obtain the molecular, statistical and graphical results.

The study of DNA sequences obtained from the GenBank, allowed us to classify 12 different genes (ORF1ab, ORF3a, N, S, ORF8, ORF10, ORF1a, ORF7a, ORF7b, ORF6, E, M) of SARS Cov 2 throughout the world The use of different software based on the determination of mutations and haplotypic diversity allows us to classify these 12 genes into three categories: the most variable genes: ORF1ab, ORF3a, N, S; the least variable genes: ORF8, ORF10, ORF1a, ORF7a, ORF7b and the non-variable genes: ORF6, E, M. The results of phylogenetic, statistical and graphical analyses of the different categories of software used show that there is a complementarity between these different bioinformatics tools.

Key words: SARS Cov2, molecular, mutations, phylogenetic, software.

Sommaire

INTRODUCTION

CHAPITRE I : RECHERCHE BIBLIOGRAPHIQUES

I.1. SARS Covid 2	1
I.1.1. Présentation	1
I.1.2. Classification	2
I.1.3. Description	2
I.1.4. Caractéristiques	3
I.1.4.1. Caractéristiques biologique	3
I.1.4.2. Caractéristiques épidémiologiques et cliniques	4
I.1.5 Transmission	5
I.1.6 Outils diagnostiques de l'infection à SARS Cov 2	5
I.1.7 Les mutations de Covid 19	6
I.1.8 Propagation dans le monde	6
I.1.9 Covid 19 en Algérie	7
I.2. Etude phylogénétique	9
I.2.1. Définition de la phylogénie	9
I.2.2. Phylogénie moléculaire	9
I.2.3. Méthodes de classification phylogénétique	10
I.2.3.1. Classification phylogénétique – cladistique	10
I.2.3.2. Arbres phylogénétiques	11
I.2.3.2.1. Les variétés d'arbres	11
I.2.3.2.2. Méthodes de reconstruction d'arbres phylogénétiques	12
I.2.3.3. Critères d'évaluation des méthodes d'inférence d'arbres phylogénétiques	12
I.2.4. Le rôle de la bio-informatique	12
I.3. Stockage de données biologiques	13

I.3.1. Banques des données	13
I.3.1.1. Les banques généralistes	13
I.3.1.2. Les banques spécialistes	14
I.3.2. Les logiciels utilisés dans la phylogénie	15
I.3.2.1. Logiciels de nettoyage	15
I.3.2.2. Logiciels d'alignement	16
I.3.3. Les méthodes d'analyses phylogénétiques.....	16
I.3.3.1. La méthode UPGMA	16
I.3.3.2. La méthode du Neighbour Joining	16
I.3.3.3. La méthode du maximum de vraisemblance	16
I.3.3.4. La méthode du maximum de parcimonie	17

CHAPITRE II : MATERIEL ET METHODES

II.1. Diagnostic de SARS Cov 2	18
II.2. Extraction des données.....	21
II.3. Traitement des séquences de la GenBank.....	25
II.3.1. FABOX	25
II.3.2. Le programme Adn Fasta format-MEGA_v1.0.....	26
II.3.3. MEGA	26
II.3.4. DnaSP	29
II.3.5. DNA Basis Analysis (DnBA).....	31
II.3.6. DAMBE	31
II.3.7. Network.....	32
II.3.7.1. Méthode 1 basée sur les haplotypes	33
II.3.7.2. Méthode 2 basée sur les séquences directement	35

CHAPITRE III : RESULTATS ET DISCUSSION

III.1. Résultats.....	36
III.1.1. GenBank	36
III.1.2. Résultats de FaBox	38

III.1.3. Résultats de programme Adn Fasta format-MEGA v1.0	38
III.1.4. MEGA.....	39
III.1.5. DnaSP.....	41
III.1.6. DnBA (DNA Bases Analyses).....	43
III.1.7. DAMBE.....	44
III.1.8. NetWork	46
III.1.8.1. Résultats de la méthode 1	47
III.1.8.2. Résultats de la méthode 2	47
III.2. Discussion.....	48
III.2.1. Comparaison général entre les logiciels utilisés	48
III.2.2. Comparaison entre les séquences de SARS Cov 2 de différents pays	50

CONCLUSION

RÉFÉRENCES BIBLIOGRAPHIQUES

BIBLIO-WEB

ANNEXE

Introduction

La bioinformatique est une discipline à multiples facettes combinant de nombreux domaines scientifiques, notamment la biologie computationnelle, les statistiques, les mathématiques, la biologie moléculaire et la génétique (**Fenstermacher, 2005**). C'est un champ de recherche où travaillent de concert biologistes, informaticiens, mathématiciens, physiciens, chimistes..., elle résout les problèmes scientifiques posés par la biologie, la bioinformatique décrit également toutes les applications informatiques résultant de ces recherches, utilisées pour les analyses génomiques, modélisation moléculaire, analyse d'image, séquençage du génome et même pour la reconstruction phylogénétique des arbres (phylogénie) (**Basseur, 2015**).

Après l'apparition du nouveau coronavirus (SARS Cov 2) qui a été détecté pour la première fois à Wuhan, en Chine, il s'est propagé rapidement depuis décembre 2019, et devenant une pandémie mondiale (**Richard et al., 2020**). Le coronavirus exprimé cliniquement par un syndrome de détresse respiratoire aiguë (SDRA), c'est un virus qui cause une pandémie mondiale, affectant des millions de personnes et mettant fin à la vie de plus de trois millions de personnes à travers le monde. Dans le but de bien connaître le virus de Cov-19 nous avons lancé ce travail qui a pour objectif d'inventorier les gènes étudiés chez ce virus ainsi que les différentes mutations détectées jusqu'à présent dans plusieurs pays, pour cela nous avons fait appel à la bioinformatique afin de réaliser des analyses phylogénétiques telles que la construction des arbres et des réseaux phylogénétiques.

Notre travail 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, description, outils de diagnostic, propagation dans le monde ; en parallèle nous avons étudié le Cov 19 en Algérie.
- ✚ Définitions des études phylogénétiques qui aident à la reconstruction phylogénétique des gènes.
- ✚ Exposition des différentes banques de données, les logiciels utilisés et les méthodes de reconstruction 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.

I.1.SARS Covid 2

I.1.1. Présentation

Les Coronavirus sont des virus à ARN simple brin non segmenté de 26 à 32 kilo bases de longueur, les membres sont nommés d'après leurs apparences en formes de couronnes au microscope électronique. Par conséquence, le 11 février 2020, le comité international pour la classification des virus a désigné le nom de ce Coronavirus comme le Coronavirus 2 du syndrome respiratoire aigu sévère (SARSCov2) (Sukhadeo et al., 2020). Le SARSCov2 apparu en décembre 2019 à Wuhan en Chine, est le troisième Coronavirus (après MERS Covid et SARS Covid) hautement pathogène à infecter les humains, similaire au Coronavirus du syndrome respiratoire aigu sévère (SARSCov) et utilise le même récepteur d'entrée cellulaire, l'enzyme de conversion de l'angiotensine 2 (ACE2) (Huihui et al., 2020).

I.1.2. Classification

La classification du Coronavirus 19 appartient à la famille Coronaviridae, sous-famille Orthocoronavirinae et genre Betacoronavirus représenté dans le tableau I.

En observant le SARS Covid 2 par microscope électronique on constate qu'il a la forme d'une couronne ; voir la figure 1.

Tableau I : Classification de Covid 19 (Zi-Wei et Shuofeng, 2020 ; Patrick et al., 2020 ; Kapikian, 2020).

Type	Virus
Domaine	Riboviria
Règne	Orthornavirae
Embranchement	Pisuviricota
Classe	Pisoniviricetes
Ordre	Nidovirales
Sous-ordre	Cornidovirineae
Famille	Coronaviridae
Sous-famille	Orthocoronavirinae
Genre	Betacoronavirus
Sous-genre	Sarbecovirus : SARS-Cov2

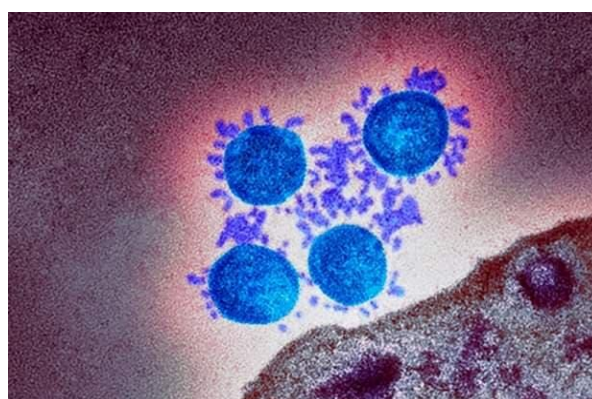


Figure 1 : Micrographie électronique à transmission couleur (TEM) d'une particule de Coronavirus SRAS Cov2 (Hopkins, 2021).

I.1.3. Description

Dans le tableau II nous avons présenté la description morphologique de SARS CoV2 ; c'est un virus à ARN de grande taille, son génome est composé d'un filament d'ARN qui porte 29981 bases codant pour 9860 acides aminés, conduisant à la production des protéines. Il code d'une part pour un grand transcrite qui sera traduit en protéines coupées par des protéases dite "non structurelles" et une partie du génome qui porte quatre gènes indépendants codant pour les protéines S,E,M,N dite "structurelles" ; voir la figure 2.

Tableau II : Description morphologique et génomique de Covid 19 (Sukhadeo et al., 2020 ; Hong et al., 2020 ; Korsia-Meffre, 2021 ; Laib, 2021).

<p>Morphologie</p>	<p>Les Cov sont enveloppés, avec un ARN simple brinde sens positif dont la taille varie entre 26 et 32 kb. Le génome du virion du SRASCov2 est d'environ 29,9 kb avec une nucléocapside enterrée à l'intérieur de la bicouche phospholipidique couverte par deux types différents de protéines de pointe, à savoir, le trimmer de glycoprotéine de pointe (S) avec protéine de membrane (M) et protéine d'enveloppe (E). Les protéines de pointe se trouvent dans tous les Cov, tandis que l'hémagglutinine-estérase (HE) ne se trouve que dans certains Cov.</p>
<p>Génome</p>	<p>L'organisation génomique du SRAS Cov 2 était cohérente avec un ARN simple brin de sens positif, qui contient à 5' une coiffe méthylée et à 3' une queue polyadénylée ; porte 29 981 bases codant pour 9 860 acides aminés, conduisant à la production de 20 protéines. Le génome est disposé dans l'ordre suivant: 5'; réplicase ORF1a/b; les gènes de structure : pointe(S), enveloppe (E), membrane(M), nucléoprotéine(N) ; et des gènes non structurelles : les protéines accessoires telles que ORF 3, 6, 7a, 7b, 8 et 9b ; 3'.</p>

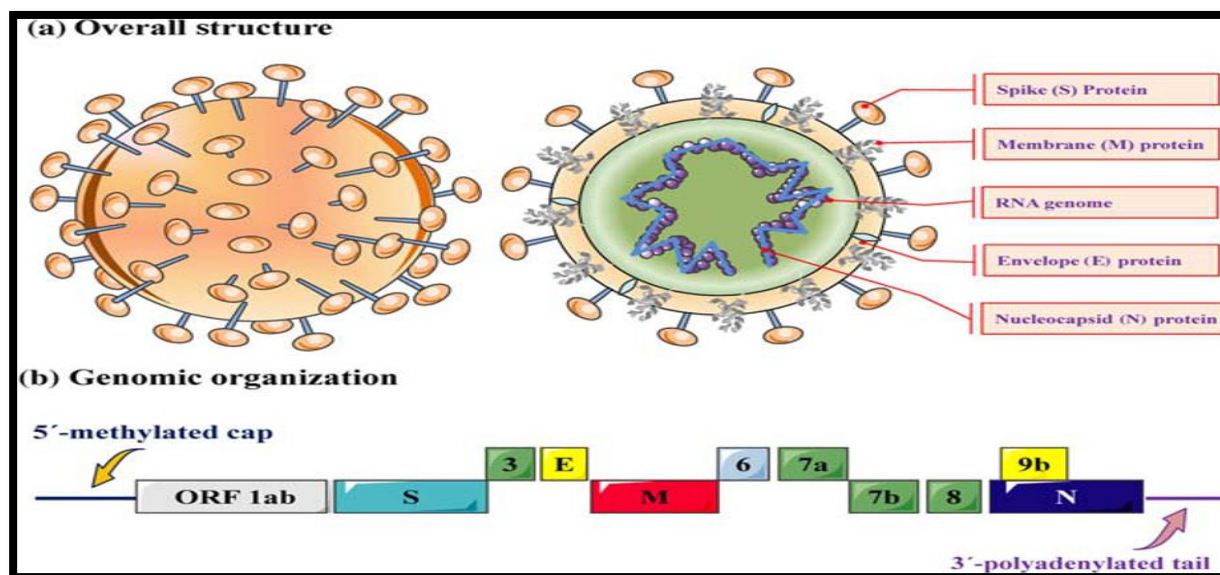


Figure 2 : Présentation de Covid19(Hong *et al.*,2020).

I.1.4. Caractéristiques

I.1.4.1. Caractéristiques biologique

Le SARSCov2 est un virus àARN de grande taille. Son génome atteint 30 000 bases. Il code d'une part pour un grand transcrit qui sera traduit en 16 protéines coupées par des protéases. Ces protéines serviront à la réplication du virus et à la formation de nouveaux brins d'ARN. À côté de ce grand transcrit, une partie du génome porte quatre gènes indépendants codant pour des protéines dont la protéine S (spike) qui, associées en trimère, donnent les images de couronne (« Corona » virus) à la surface du virus. C'est cette même protéine qui se lie au récepteur dans le tissu pulmonaire. Les autres protéines correspondent aux protéines d'enveloppe (M, E et N). Le SARSCov2 commence son cycle lorsque sa protéine S se lie au récepteur cellulaire. Ce récepteur correspond à l'enzyme de conversion de l'angiotensine 2 (ACE2), que l'on trouve à la surface de pneumocytes, de cellules endothéliales mais également de cellules de l'endocarde, du foie et de façon importante dans l'intestin (entérocytes, colonocytes et cholangiocytes). Le SARSCov2 peut ainsi se lier à plusieurs tissus chez un même individu. Son récepteur est assez inattendu car cet enzyme, ACE2, participe au clivage et à la dégradation d'un pro-enzyme, l'angiotensine 1, qui régule la tension artérielle. Plusieurs études sont en cours pour comprendre les liens entre la gravité de la maladie Covid-19, l'hypertension artérielle (HTA) et les traitements de type inhibiteur de l'enzyme de conversion de l'angiotensine (IEC) ou antagonistes des récepteurs de

l'angiotensine 2 (ARA 2). Après la liaison au récepteur, le changement de conformation de la protéine S facilite la fusion de l'enveloppe virale avec la membrane cellulaire par la voie endosomale. SARSCov2 libère son ARN dans la cellule hôte. L'ARN de son génome est ensuite traduit en protéines qui sont secondairement clivées pour former les protéines structurales et de multiplication virale. Les protéines virales et l'ARN du génome sont ensuite assemblées en virions dans le réticulum endoplasmique et l'appareil de golgi puis bourgeonnent dans la lumière du RE (Purva *et al.*, 2020). (Figure3) .

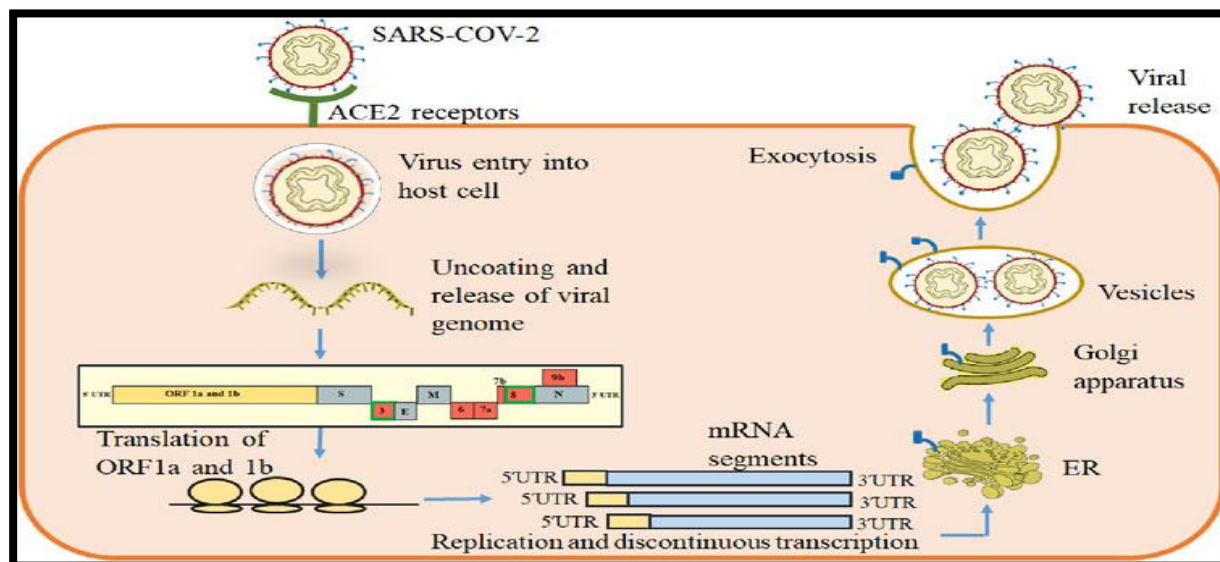


Figure 3 :Cycle virale de SARS Covid 19 (Purva *et al.*, 2020).

I.1.4.2. Caractéristiques épidémiologiques et cliniques

Des enquêtes épidémiologiques ont montré que les caractéristiques cliniques de l'infection par le SRASCov2 sont similaires à celles du SARSCov et se caractérisent par de la fièvre (> 37,3 C), une toux sèche, une dyspnée ou un essoufflement chez la plupart des patients, alors que des symptômes cliniques tels que diarrhée, maux de gorge, douleurs musculaires, maux de tête et vomissements ont également été rapportés chez une minorité de patients. Les patients confirmés ont également développé un syndrome de détresse respiratoire aigüe, tandis qu'une maladie grave peut se traduire par une défaillance des fonctions respiratoire et pulmonaire, voire un dysfonctionnement de plusieurs organes et un choc septique, qui nécessitent une oxygénation extracorporelle de la membrane (ECMO) et un soutien en soins intensifs. Les principales caractéristiques pathologiques du poumon comprennent l'alvéolite pulmonaire et la bronchiolite avec prolifération des cellules épithéliales, desquamation, métaplasie squameuse et production de mucus et de liquide œdémateux(Hong *et al.*, 2020) .

I.1.5. Transmission

Le SARSCov 2 peut se propager à la fois par :

-**Des moyens directs** (gouttelettes et transmission interhumaine).

-**Contact indirect** (objets contaminés et contagion aérienne).

Pendant ce temps, les équipements de protection individuelle (EPI) pourraient également être à l'origine d'infections aéroportées. La propagation de personne à personne du SARSCov2 est censée se produire principalement via des gouttelettes respiratoires, lorsqu'un patient tousse, éternue ou même parle ou chante. Les gouttelettes ne peuvent généralement pas traverser plus de six pieds (près de deux mètres) et restent dans l'air pendant un temps limité. Le SARSCov2 peut être suspendu dans l'air jusqu'à trois heures. Par conséquent, l'isolement aérien, la ventilation de la pièce et l'application appropriée de désinfectant (en particulier dans les toilettes) peuvent limiter la propagation du virus par aérosol. Le Covid19 peut survenir si une personne touche une surface contaminée par le SARSCov2, puis les mains entrent en contact direct avec des muqueuses telles que les yeux, le nez ou la bouche. Ainsi, il est recommandé de se laver les mains avec de l'eau et du savon ou avec un désinfectant pour les mains. La propagation du SARSCov2 à partir d'individus asymptomatiques (ou d'individus pendant la période d'incubation), sans aucun fadings, a également été signalé. Par conséquent, il existe un besoin pour des améliorations des méthodes de diagnostic rapides et sensibles pour détecter les individus infectés (Melika et al., 2020).

I.1.6. Outils diagnostiques de l'infection à SARSCov2

Le diagnostic de certitude repose sur l'identification du virus par RT-PCR effectuée sur prélèvements respiratoires. On prélève un échantillon sur les parties du corps où le virus responsable de la Covid-19 s'accumule, comme le nez ou la gorge. On traite l'échantillon avec plusieurs solutions chimiques pour le débarrasser de certaines substances, notamment les protéines et les graisses, et extraire uniquement l'ARN qu'il contient. Cependant, devant l'accessibilité à la RT-PCR parfois limitée et la fréquence élevée de faux négatifs de cette technique (environ 30 %), le diagnostic de COVID-19 peut être porté par l'association de signes cliniques évocateurs et d'une image scénographique compatible. L'utilisation de la sérologie est en cours d'évaluation (Plaçais et Richier, 2020 ; Jawerth, 2020) (Figure 4).

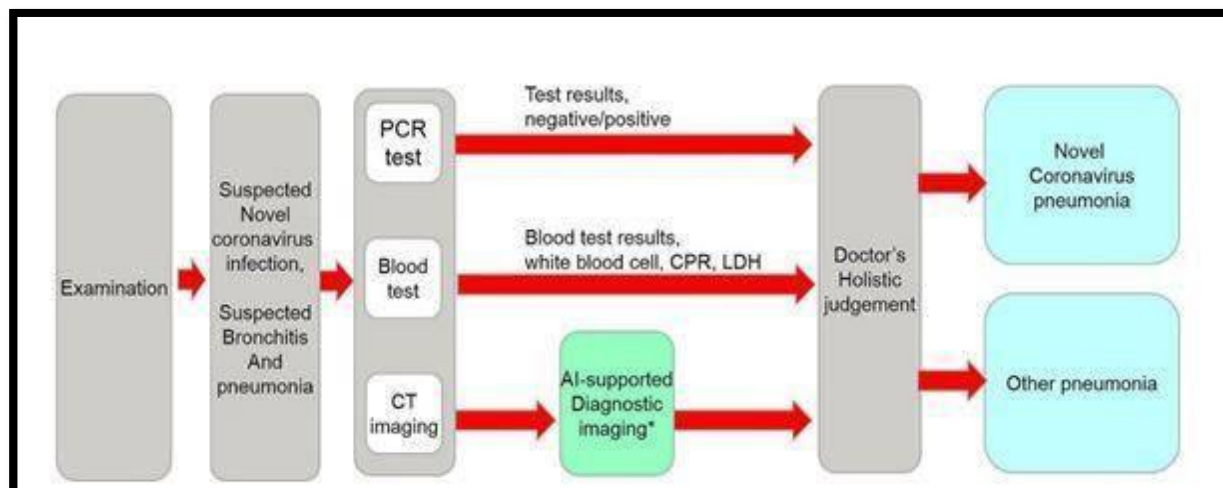


Figure 4 : Les approches diagnostiques et thérapeutiques de covid19(Chanthadavong, 2020).

I.1.7. Les mutations de Covid 19

Depuis plusieurs semaines, l'actualité COVID-19 est emportée dans une tourmente phylogénétique où tourbillonnent mutations, lignées et variants, le plus souvent affublés de nationalités : anglais, sud-africain, brésilien, etc.

Comme pour tous les virus à ARN, la réplication de SARS-CoV-2 n'est pas très fidèle et de nombreuses mutations apparaissent rapidement. Ces mutations sont essentiellement des substitutions (Korsia-Meffre, 2021).

I.1.8.Propagation dans le monde

Selon Moyou (2021) et le site officiel de l'organisation mondiale de la santé (OMS, <https://www.who.int/fr/>) contenant les statistiques de (Covid19), nous avons les résultats suivants : Sur un total de 183 millions de personnes infectées dans le monde entier, 76.5 millions ont été guéries (juillet 2021). Le tableau III suivant représente les statistiques.

Tableau III : Statistiques de Covid 19 dans le monde et dans quelques pays (Moyou,2021).

Lieux	Cas	Guérisons	Décès
Monde entier	183 M	76,5 M	3,97 M
États-Unis	33,7 M	33 698 562	605 k
Brésil	18,7M	16,5M	522 k
France	5,78M	600 064	111 k
Russie	5,5 M	4.98 M	134k

Nigeria	168 k	164 k	2121
Kenya	185 k	127 k	3651
Maroc	533 k	519 k	9307
Tunisie	433 k	359 k	15 179
Royaume-uni	4,86 M	4 222 263	128 k

I.1.9. Covid 19 en Algérie

L'Algérie est parmi les premiers pays qui ont déclaré sur la page officielle de l'institut pasteur algérien (<https://www.pasteur.dz/fr/>), la présence des cas confirmés de Covid 19. Selon les journaux officiels (**El Watan** et **le Figaro**) et les déclarations officielles du ministère de la santé algérienne le nombre total de contaminants atteint 141 k. le tableau suivant représente l'état de ce virus en Algérie jusqu'à la réalisation de notre travail (juillet 2021).

Tableau IV : SARS Covid 2 en Algérie (Kourta, 2021 ; Fourar, 2020 ; Laib, 2021).

Historique	En Algérie, le SARS Covid 2 se propage à partir du 25 février 2020 lorsqu'un ressortissant italien est testé positif au SARS-CoV-2. À partir du 1er mars 2020, un foyer de contagion se forme dans la wilaya de Blida , seize membres d'une même famille ont été contaminés par le Coronavirus lors d'une fête de mariage à la suite de contacts avec des ressortissants algériens en France. Progressivement, l'épidémie se propage pour toucher toutes les wilayas algériennes.		
Statistique (Juillet 2021)	Cas	Guérisons	Décès
	141 K	97701	3 735
Les souches mutantes déclarées	<p>Souche britannique : l'Algérie a enregistré 78 cas de cette souche, détectés le 19 février 2021 pour la première fois chez un membre du personnel de santé de l'EHS de Psychiatrie de Chéraga et chez un émigré venant de France pour l'enterrement de son père.</p> <p>Souche nigérienne : Son apparition la première fois est durant la dernière semaine de février 2021. Six cas ont été</p>		

	<p>découverts dans la wilaya de Tebessa, trois autres à Alger, deux dans la wilaya de In Salah, et les deux derniers à Souk Ahras et Ouargla. Au total, on compte treize de cette deuxième variante africaine dans le pays. En avril 2021, le nombre de contaminants a augmenté à 129 cas.</p> <p>Souche indienne : Souche indienne Six cas du variant indien (B.1.617) ont été détectés pour la première fois, dans la wilaya de Tipaza le 3 mai 2021.</p>	
Les méthodes d'analyse	RT-PCR	
	Sérologie et image scénographique	
Quelques laboratoires appliquant la RT-PCR	Alger Centre Institut Pasteur	
	Alger Épsilon	
	Bejaia Ait_Bachir	
	Ghardaïa Ibn_Rochd	
	Tipaza	Khoury
	Oran	Belarbi
	Meftah (Blida)	Sifi
	Kouba (Alger)	Biotop
Traitement	<p>L'Algérie a décidé fin mars 2020 de soigner les patients atteints du nouveau Coronavirus avec un double traitement de chloroquine et l'azithromycine qui est un antibiotique. Depuis, plus de 15 000 personnes contaminées ou suspectées de l'être ont reçu un traitement à base d'hydroxychloroquine.</p>	
Vaccination	<p>La campagne de vaccination a commencé le samedi 30 janvier 2021 en Algérie, au lendemain de l'arrivée des premières doses du vaccin russe Spoutnik V. La ville de</p>	

Blida fut la première étape de cette campagne, le nombre totale de doses jusqu'à maintenant est **75 000 doses**.

Types de vaccins : Vaccin russe Spoutnik V

Vaccin anglo-suédois Astra Zeneka

Vaccin chinois Sinopharm

I.2. Etude phylogénétique

I.2.1. Définition de la phylogénie

La phylogénie est la branche de la biologie qui étudie les relations évolutives entre les êtres vivants. Ces relations sont représentées à l'aide d'arbres phylogénétiques, construits à partir de données moléculaires (aujourd'hui) ou morphologiques (avant l'avènement de la biologie moléculaire)(Taguan,2015).

I.2.2. Phylogénie moléculaire

La phylogénie moléculaire est basée sur l'analyse comparative des séquences d'ADN ou de protéines. Elle complète les méthodes traditionnelles de classification basées sur l'observation des caractères morphologiques et anatomiques.

La phylogénie moléculaire utilise les gènes des organismes vivants pour élaborer des arbres phylogénétiques en comparant les caractères moléculaires (séquences nucléotidiques ou protéiques).

La phylogénie moléculaire essaie de retracer l'accumulation des mutations dans les génomes au cours de l'évolution des espèces. Elle permet de reconstruire l'histoire évolutive des organismes. Les espèces ont des génomes d'autant plus proches qu'elles ont divergé récemment depuis leur ancêtre commun

Les régions du génome utilisé pour la phylogénie moléculaire sont appelées marqueurs moléculaires. Elles doivent remplir les conditions suivantes:

- 1/ Elles doivent être ubiquistes, c'est-à-dire présentes chez tous les taxons étudiés.
- 2/ Elles doivent contenir une région variable spécifique de chaque espèce et encadrée par deux séquences conservées (**Baaziz, 2017**).

Les marqueurs d'ADN utilisés pour évaluer la diversité génétique sont :

✚ **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 (Sunnucks, 2001).

✚ **Marqueurs d'ADN mitochondrial** : Les polymorphismes d'ADN mitochondrial (ADNmt) ont été largement utilisés lors des analyses de la diversité phylogénétique et génétique. L'ADNmt haploïde transporté par les mitochondries du cytoplasme cellulaire, possède un mode maternel d'hérédité (les animaux héritent l'ADNmt de leurs mères et non de leurs pères) et un taux de mutation élevé; il ne se recombine pas. Ces caractéristiques consentent aux biologistes de reconstruire les relations évolutives intra et interraciales par l'évaluation des modèles de mutation de l'ADNmt. Les marqueurs d'ADNmt peuvent également fournir un moyen rapide pour détecter l'hybridation entre les espèces et les sous-espèces d'animaux d'élevage (Nijman *et al.*, 2003).

I.2.3. Méthodes de classification phylogénétique

I.2.3.1. Classification phylogénétique – cladistique

Les temps modernes ont vu l'apparition de la classification phylogénétique, un système de classification des êtres vivants (Hennig, 1950). Elle se base sur les rapports de proximité évolutive entre espèces. Cette classification se veut plus objective, seuls les caractères empiriquement observables et propres au rang d'espèces sont retenus comme uniques témoins de l'héritage ancestral de chaque espèce (Lecointre et Guyader, 2002).

Cette objectivité est pourtant fortement liée à la vision évolutive moderne. La représentation schématique est donnée par un cladogramme, qui est un arbre phylogénétique non enraciné, contenant des nœuds et des feuilles (appelés taxons). Les groupes qui incluent un ancêtre commun et tous ses descendants sont appelés groupes monophylétiques (Allaby, 2009).

I.2.3.2. Arbres phylogénétiques L'arbre phylogénétique est une représentation graphique de la phylogenèse d'un groupe de taxons (séquences).

- ✓ Arbre non enraciné : graphe connexe non cyclique. Il n'y a qu'un seul et unique chemin pour passer d'un sommet à l'autre (Figure 5).

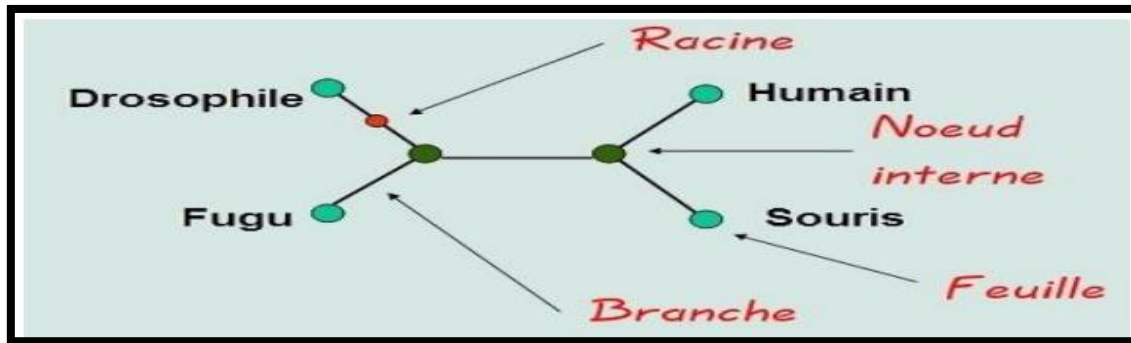


Figure 5 : Arbre phylogénétique non enraciné (Benhamdi, 2020).

- ✓ Arbre enraciné : possède une contrainte supplémentaire par rapport au précédent. Présence de liens orientés depuis une origine ou un ancêtre (Figure 6).

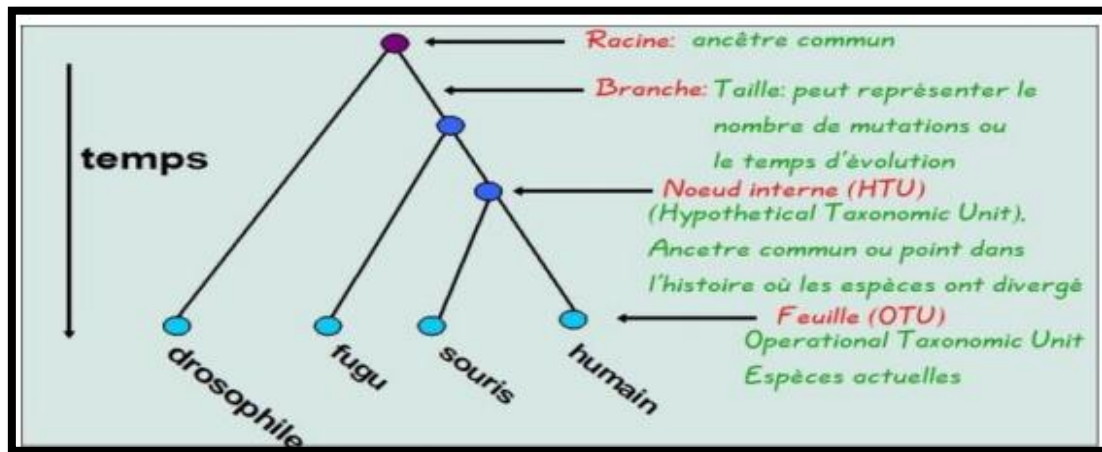


Figure 6 : Arbre phylogénétique enraciné (Benhamdi, 2020).

I.2.3.2.1 Les variétés d'arbres

Dendrogramme : Arbre exprimant les liens entre taxons sous la forme d'une succession de branchements. Il existe plusieurs types de dendrogrammes selon les méthodes avec lesquelles ils ont été construits.

Cladogramme : Dendrogramme exprimant les relations phylogénétiques (de parenté) entre taxons et construit à partir d'une analyse cladistique. Chacun des points de branchements ou nœuds, est défini par une ou plusieurs synapomorphies.

Phénogramme : Un dendrogramme obtenu par méthodes de distance où les relations entre taxa expriment des degrés de similitude globale;

Phylogramme : C'est un cladogramme dont la longueur des branches est proportionnelle au nombre de changements évolutifs.

Arbres additifs: Longueur des branches proportionnelle au nombre de changements évolutifs (Benhamdi, 2020).

I.2.3.2 Méthodes de reconstruction d'arbres phylogénétiques

Il existe deux grands types de méthodes permettant la reconstruction d'arbres phylogénétiques :

- les méthodes basées sur les mesures de distances entre séquences prises deux à deux, c'est à dire le nombre de substitutions de nucléotides ou d'acides aminés entre

Ces deux séquences, ex : UPGMA et NJ

- les méthodes basées sur les caractères qui s'intéressent au nombre de mutations (substitutions /insertions /délétions) qui affectent chacun des sites (positions) de la séquence, ex : la parcimonie (Benhamdi, 2020).

I.2.3.3 Critères d'évaluation des méthodes d'inférence d'arbres phylogénétiques

Plusieurs critères de qualité sont utilisés :

- **Efficacité** (complexité asymptotique, espace mémoire).
- **Puissance** (bonne utilisation des données sans perte d'information).
- **Consistance** (convergence vers une même solution avec des données différentes et même structure du problème).
- **Robustesse** (bons résultats en dépit de la violation du modèle initial choisi).
- **Falsifiabilité** (possibilité de reconnaître quand les résultats de la méthode sont inappropriés) (Huelsenbeck, 1995 ; Moret et al., 2002).

I.2.4. Le rôle de la bio-informatique

L'élaboration de technologies à haut débit serait inutile si l'on ne disposait pas des 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. 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. Stockage de données biologiques

I.3.1. Banques des données

Les fichiers contenant l'information biologique sous la forme de séquences est l'élément central autour duquel les banques de données se sont constituées à l'origine.

On peut distinguer :

- ✚ **Les bases de données généralistes** : Elles correspondent à une collecte des données la plus exhaustive possible et qui offrent un ensemble plutôt hétérogène d'informations.
- ✚ **Les bases de données spécialisées** : Elles correspondent à des données plus homogènes établies autour d'une thématique et qui offrent une valeur ajoutée.

Il existe un grand nombre de bases de données d'intérêt biologique. Le panorama de ces milliers de bases de données biologiques nécessite cependant un préalable qui s'appuie sur une forme de "sagesse" :

- La maintenance et la mise à jour actives de bases de données biologiques publiques sur le Web demandent beaucoup temps et sont coûteuses.
- Sans soutien institutionnel ou plan de viabilité financière, la plupart des bases de données créées en tant que résultats de projets de recherche meurent ou sont archivées dans un délai de 10 à 15 ans (Benalioua B., 2011).

I.3.1.1. Les bases de données généralistes

Les bases de données généralistes sont indispensables à la communauté scientifique car elles regroupent des données et des résultats essentiels dont certains ne sont plus reproduits dans la littérature scientifique.

- **NCBI** "National Center for Biotechnology Information"
- **EBI** "European Bioinformatics Institute"
- **Uniprot**
- **PDB** "Protein Data Bank"

Dans le cadre de l'analyse des séquences, par exemple, le fait que la majorité des séquences connues soit réunie en un seul ensemble est un élément fondamental pour la recherche de

similitudes avec une nouvelle séquence. D'autre part, la grande diversité d'organismes qui y est représentée permet d'aborder des analyses de type évolutif.

La principale mission des bases de données généralistes est de rendre publiques les séquences et tout autre type d'information. Cette notion de mise à la disposition du public a été capitale dans le cas par exemple de la diffusion des résultats de l'humain. On y trouve également de l'information qui accompagne les séquences (annotations, bibliographie, ...) et une expertise biologique directement liées aux séquences traitées (**Benalioua, 2011**).

Faiblesses des banques généralistes

✚ Hétérogénéité dans la nature des séquences :

ADN nucléaire ou mitochondrial, ARN (t, r, m, ...), génome.

✚ Variabilité de l'état des connaissances :

Caractérisation biologique beaucoup plus lente que le séquençage.

✚ Erreurs dans les séquences :

Liées à l'origine du fragment, à la technologie, à la méthodologie.

✚ Biais d'échantillonnage :

Desespèces. Desgènes. Redondance des données.

Pour cela il ya le besoin de création de banques spécialisées (**Corre, 2013**).

I.3.1.2. Les bases de données spécialisées

Pour des besoins spécifiques liés à l'activité d'un groupe de personnes, ou encore par compilations bibliographiques, de nombreuses bases de données spécifiques ont été créées au sein des laboratoires. Certaines sont inconnues ou mal connues et attendent qu'on les exploite davantage. Les bases de données spécialisées sont d'intérêt divers et la masse des données qu'elles contiennent peut varier d'une base à une autre. Ces bases correspondent à des améliorations ou à des regroupements par rapport aux données issues des bases généralistes (**Jaspard, 2021**).

Exemples de bases de données spécialisées :

-Late Embryogenesis Abundant Proteins database LEAPdb (Hunault et Jaspard, 2010): Cette base de données contient un grand nombre d'informations sur les protéines LEA impliqués dans la tolérance à de nombreux stress, notamment la déshydratation et le froid. Pour l'instant, on les a mises en évidence principalement chez les plantes.

-**Small HeatShockProteinsdatabase_sHSPdb** (Jaspard et Hunault, 2016) : Cette base de données contient un grand nombre d'informations structurales sur les protéines de choc thermique de faible masse molaire.

-**RESIDdatabase** : Base de données sur les acides aminés peu fréquents (sous-partie de la base de données PIR) (Jaspard, 2021).

I.3.2. Les logiciels utilisés dans phylogénie

La phylogénie moléculaire a pour but de reconstruire les relations de parenté entre des séquences de nucléotides ou d'acides aminés. On peut ainsi étudier les relations de parenté entre les espèces qui les portent mais, aussi, l'évolution du génome. En particulier, pour chaque famille multigénique, on peut déterminer l'importance relative des événements de duplications et de transferts horizontaux de gènes. La fiabilité des méthodes de reconstruction phylogénétique repose sur la compréhension des mécanismes d'évolution des séquences, un domaine qui a beaucoup progressé ces dernières années. Cela a abouti à une vision sans cesse plus correcte de l'arbre universel du vivant. L'étude des contraintes fonctionnelles agissant sur les protéines bénéficie de ces avancées. En particulier, la détection, dans une protéine, des positions qui sont soumises à une sélection darwinienne est devenue assez performante, permettant de prédire les substitutions à l'origine d'un changement de fonction et donc de guider les études expérimentales. Dans cette section, nous allons décrire quelques grandes catégories de logiciels utilisés pour les analyses phylogénétiques (Philippe et al., 2002).

I.3.2.1. Logiciels de nettoyage

Il existe des logiciels de nettoyage de l'alignement, leur travail c'est d'éliminer les régions divergentes ou peu informatives, parmi ces logiciels :

- **Sequencher** : Assemblés et alignés plusieurs séquences d'ADN contiguës relativement courtes afin de créer des séquences plus longues.
- **Bio Edit** : Pour des études moléculaires de différents organismes tels que des génomes de virus (Ron et al., 2005 ; Chen et al., 2006).

I.3.2.2. Logiciels d'alignement

Un alignement de deux séquences (appelé souvent 'Alignement deux à deux') est une mise en correspondance entre les résidus avec une possible insertion des espaces (gaps) afin

d'obtenir des séquences de longueur égales. Toutes les correspondances sont autorisées à condition que l'ordre des résidus soit respecté.

Trois situations sont possibles pour une position donnée de l'alignement :

- Les caractères sont les mêmes : Identité.
- Les caractères ne sont pas les mêmes : Substitution
- L'une des positions est un gap (espace) : Insertion/Délétion.

Parmi les logiciels qui font l'alignement on a : Blast, Clustal W et Exonerate (Yaeghobi, 2017).

I.3.3. Les méthodes d'analyses phylogénétique

I.3.3.1. La méthode UPGMA (Unweighted Pair Group Method with Arithmetic Mean) :

Est une méthode dite de distance, c'est-à-dire une méthode basée sur les similarités entre paires de séquences. Elle a vite été délaissée au profit de sa cousine (NJ) qui est plus adaptée aux études phylogéniques moléculaires (Yoann, 2012).

I.3.3.2. La méthode du NeighbourJoining (NeighborJoining- NJ) : C'est aussi une méthode de distance, elle a l'avantage d'être vraiment rapide. En général, elle est utilisée pour faire des arbres de plusieurs milliers de séquences (Yoann, 2012).

I.3.3.3. La méthode du maximum de vraisemblance (Maximum Likelihood- ML) :

C'est une méthode dite de caractère(s), elle repose sur un ou plusieurs caractères à étudier. Il s'agit d'une méthode probabiliste qui nécessite un modèle d'évolution. Le choix de ce modèle est crucial pour la qualité de l'arbre obtenu. On dit qu'il convient de l'utiliser à partir du moment où le nombre de caractères analysés est supérieur à la moitié du nombre de séquences analysées, sinon la reconstruction est considérée comme incorrecte. Elle est souvent décrite comme étant la meilleure méthode, c'est-à-dire la plus efficace pour trouver l'arbre le plus proche de la réalité. Son désavantage se situe au niveau des temps de calculs qui sont extrêmement longs, il est arrivé d'avoir des jobs tournant sur le cluster pendant plusieurs semaines pour des fichiers contenant plusieurs centaines de séquences (Yoann, 2012).

I.3.3.4. La méthode du maximum de parcimonie (Maximum Parsimony) : Elle est très appréciée car rapide en temps de calcul, mais pas aussi précise que sa cousine (ML). Comme souvent donc, on gagne du temps de calcul mais on perd de la précision.

- ✚ D'autres méthodes apparaissent régulièrement comme celles basées sur de l'apprentissage artificiel (via un réseau de neurones comme ici ou là par exemple) ou

encore de l'inférence Bayésienne comme dans le logiciel MrBayes qui est très utilisé. Néanmoins, les méthodes principalement utilisées restent la NJ et la ML (Yoann., 2012).

Cette étude est réalisée au niveau du laboratoire de génétique des populations et biologie de conservation des populations animales de l'Université des Sciences et de la Technologie Houari Boumediene (USTHB) d'Alger.

Ce mémoire reposant sur l'inventaire et traitement par des outils de bioinformatique des gènes identifiés chez le coronavirus, exprimés cliniquement par un syndrome de détresse respiratoire aigüe (SDRA) ; dans ce chapitre, nous présenterons quelques fonctions des logiciels utilisés dans notre travail.

II.1. Diagnostique de SARS Cov 2

La RT-PCR est la technique de référence pour le diagnostic de COVID-19, mais en parallèle de la RT-PCR ils existent d'autres techniques, de diagnostic rapide qui sont :

- **Tests immunologiques ou tests sérologiques** : Les tests immunologiques permettent de mesurer les anticorps (IgM et IgG circulants) de patients atteints de SARS Cov 2. On distingue les tests dit tests ELISA et les tests immunochromatographiques. Ces derniers incluent les tests rapides de détection d'anticorps et les tests rapides de détection d'antigène (**Gala et al., 2020**).
- **Les tests rapides immunochromatographiques sur bandelette de nitrocellulose (aussi appelés communément « Lateral Flow Assay » ou « LFA »)** : Ce sont des tests rapides qui permettent d'établir un diagnostic en quelques minutes (moins de 15 minutes). Une réaction chimique, enzymatique ou immunologique fait apparaître une coloration particulière permettant d'interpréter immédiatement le résultat (**Gala et al., 2020**).
- **Les tests rapides de détection des antigènes du virus SARS Cov2** : Ce test « COVID-19 Ag Respi-Strip diagnostic assay » est un test immunochromatographique développé pour la détection rapide de l'antigène Cov 2 du SRAS sur un échantillon nasopharyngé. Il a été mis au point pour identifier les patients Cov 19 positifs (**Gala et al., 2020**).
- **Les scanners thoraciques** : C'est un examen d'imagerie pulmonaire pour les patients présentant des signes de mauvaise tolérance respiratoire (**Hoppenot, 2020**).

La figure 7 suivante représente les principaux outils de diagnostics et thérapeutiques de SARS Cov2.

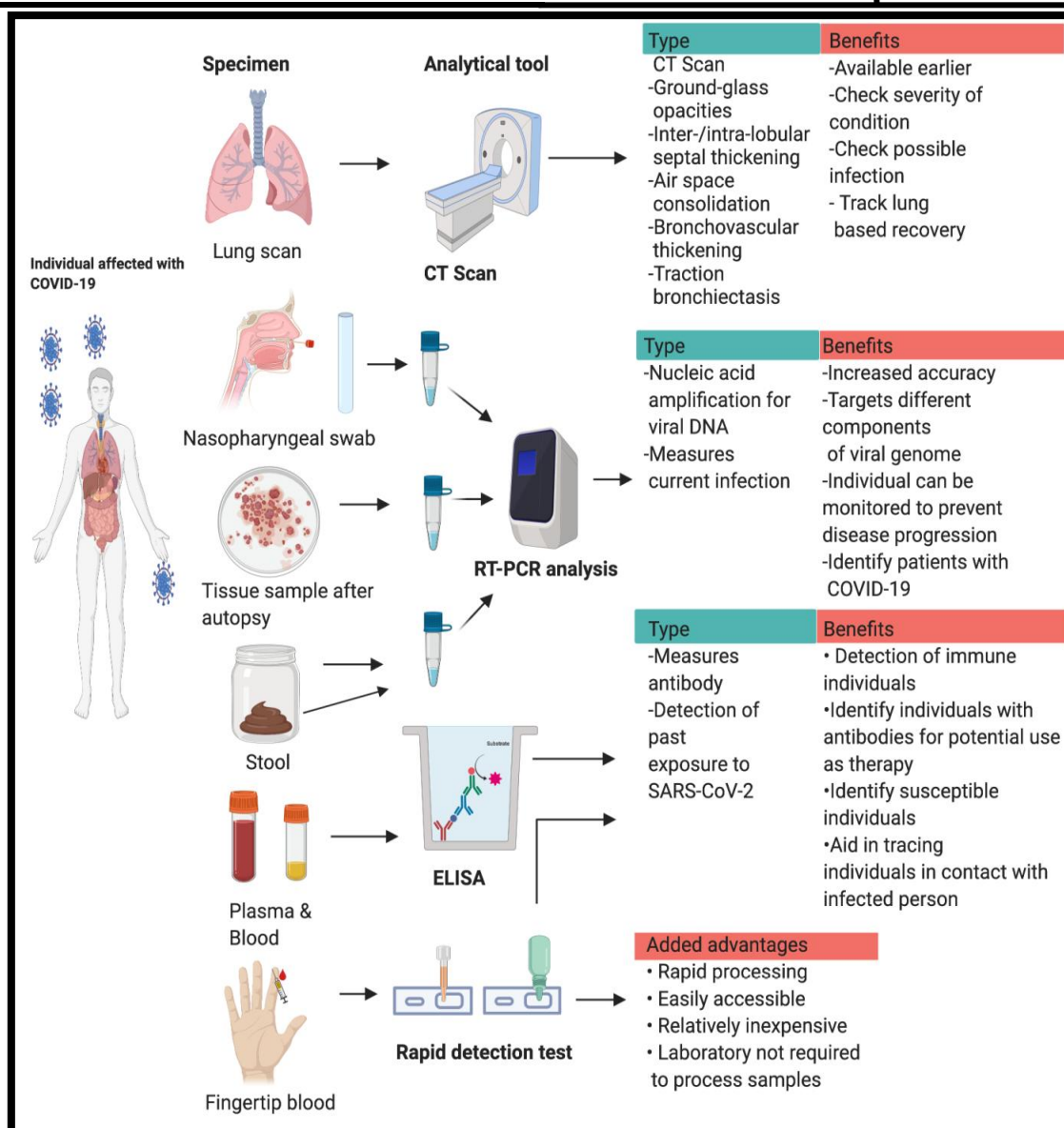


Figure 7 : Approches diagnostiques et thérapeutiques de Covid19 (Mahalaxmi et al., 2020).

Selon la formation nationale sur le COVID-19 organisée par le ministère de la santé, de la population et de la réforme hospitalière direction de la formation le 29 avril 2021 sur ZOOM présenté par : Dr. BENOUAR Salem maître assistant. Laboratoire central CHU BLIDA ; Pr. ABDI Samia professeur en biochimie, médecin chef laboratoire CHU BLIDA ; Mme Khadidja. AOUES inspectrice INFSPM BLIDA ; Dr AZI. Ourida maître assistante en Biochimie EPHTPAZA, que nous avons suivi ; l'Algérie applique des protocoles sanitaires très strictes dans les laboratoires de diagnostic de COVID-19 ; les figures 8 et 9 indiquent le matériel utilisé et l'équipement de la protection individuelle (Aoues, 2020). Ainsi l'Algérie applique des techniques de diagnostic bien étudiées entre autre la technique de la RT-PCR ou elle adapte dans le protocole de la technique comme marquer moléculaire ou gène cible de

diagnostic de SARS Cov 2 des gènes bien étudiés qui sont les gènes : N, E, et le RDRP (une partie de ORF1ab).



Figure 8 : Equipement de laboratoire de biologie moléculaire utilisé pour l'analyse de RT-PCR de SARS Cov 2 (Aoues, 2020).



Figure 9 : Equipement de protection individuelle dans laboratoire d'analyse de SARS Cov 2 (Aoues, 2020).

D'après **Plac et al. (2020)** et **Jawerth (2020)** les étapes d'analyse de SARS Cov 2 par RT PCR sont :

- Prélèvements d'un échantillon sur les parties du corps où le virus responsable de la Cov 19 s'accumule, comme le nez ou la gorge.
- Traitement d'échantillon avec plusieurs solutions chimiques pour le débarrasser de certaines substances, notamment les protéines et les graisses, et extraire uniquement l'ARN qu'il contient.
- L'ARN est alors converti en ADN, lors de la transcription inverse, grâce à une enzyme spécifique.
- L'ajout de courts fragments d'ADN complémentaires de certaines séquences de l'ADN viral transcrit. Si le virus est présent dans l'échantillon, ces fragments s'attachent aux séquences d'ADN viral cibles. Certains des fragments d'ADN ajoutés servent uniquement à construire de nouveaux brins d'ADN lors de l'amplification, tandis que les autres servent aussi au marquage des brins qui permettront de détecter le virus.
- Le mélange est ensuite placé dans un appareil de RT-PCR, où il est chauffé et refroidi suivant des cycles qui déclenchent des réactions chimiques permettant d'obtenir de nouvelles copies, identiques, des séquences d'ADN viral cibles.
- Les cycles se répètent de nombreuses fois pour continuer à copier ces séquences. À chaque cycle, la quantité double.
- Fixation des marqueurs sur les brins d'ADN et émettent une fluorescence qui est mesurée par l'ordinateur de l'appareil. Les résultats s'affichent en temps réel à l'écran. Lorsque le niveau de fluorescence dépasse un certain seuil, la présence du virus est confirmée.

Les séquences obtenues par différents laboratoires de différents pays vont être étudiées et classées par des chercheurs scientifiques puis soumises dans des banques de données moléculaires entre autre la GenBank afin de faciliter les recherches scientifiques.

II.2. Extraction des données

Nous avons choisi la GenBank pour l'extraction des données.

La **GenBank** est une base de données de l'ADN, comprenant toutes les séquences de nucléotides publiquement disponibles et leur traduction en protéines. Cette base de données américaine en libre accès, a été créée au centre national pour l'information biotechnologique (NCBI) dans le cadre de la collaboration internationale sur le séquençage

des nucléotides (INSDC selon le sigle anglais). La GenBank se construit soit par des dépôts directs en provenance de laboratoires, soit des dépôts en masse des centres de séquençage à grande échelle (Benson *et al.*, 2008).

Toutes les séquences de SARS Cov 2 étudiées dans ce travail ont été téléchargées à partir de la GenBank, nous passons par plusieurs étapes distinctes :

1. Construire l'URL suivant : <https://www.ncbi.nlm.nih.gov/genbank/>
2. Choisir recherche par taxonomie sur la source NCBI.
3. Ecrire le nom de notre genre « SARS Cov 2 » (Figure 10).

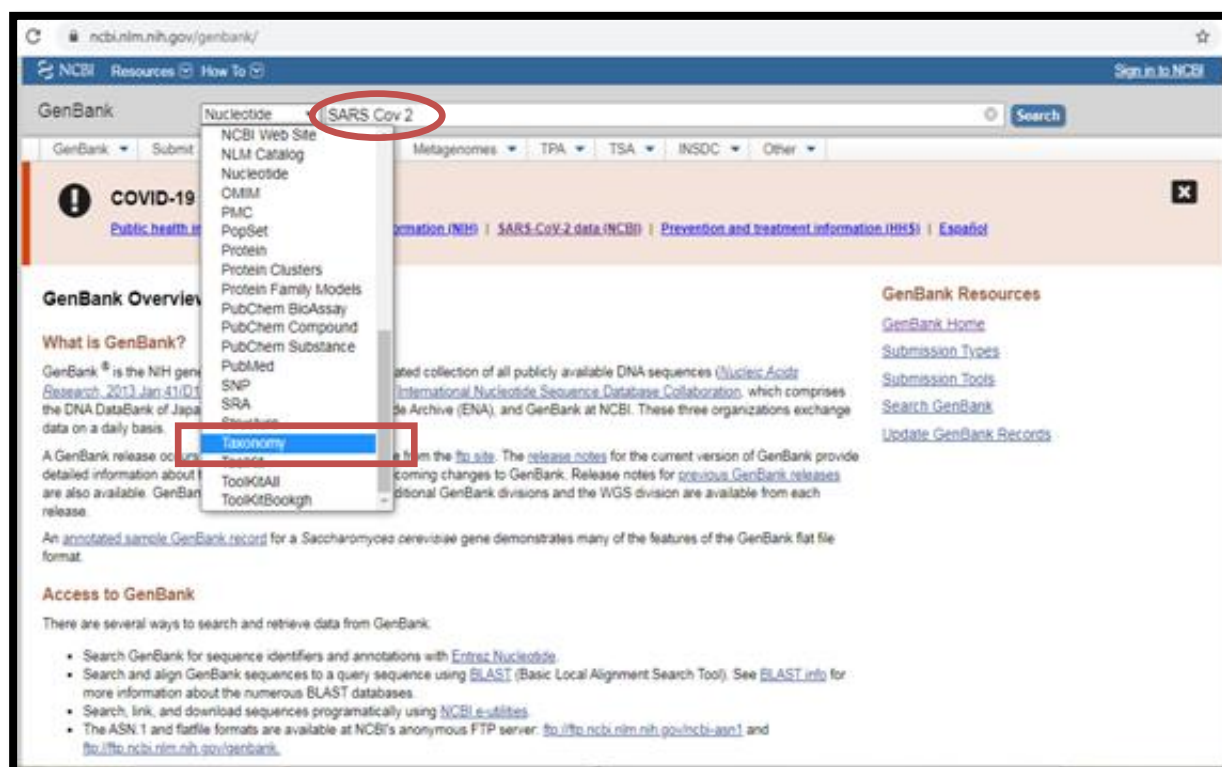


Figure 10: Les étapes initiales pour récupérer les séquences de la GenBank.

4. En entrant sur **Severe acute respiratory syndrome coronavirus 2** (Figure 11) puis sur le nombre de nucléotides (+de 498,626) dans le tableau, comme présenté dans la figure 12, nous obtenons toutes les séquences étudiées par différents laboratoires de différents pays ; la figure 13 présente quelques résultats de ces séquences, montrés par des flèches.

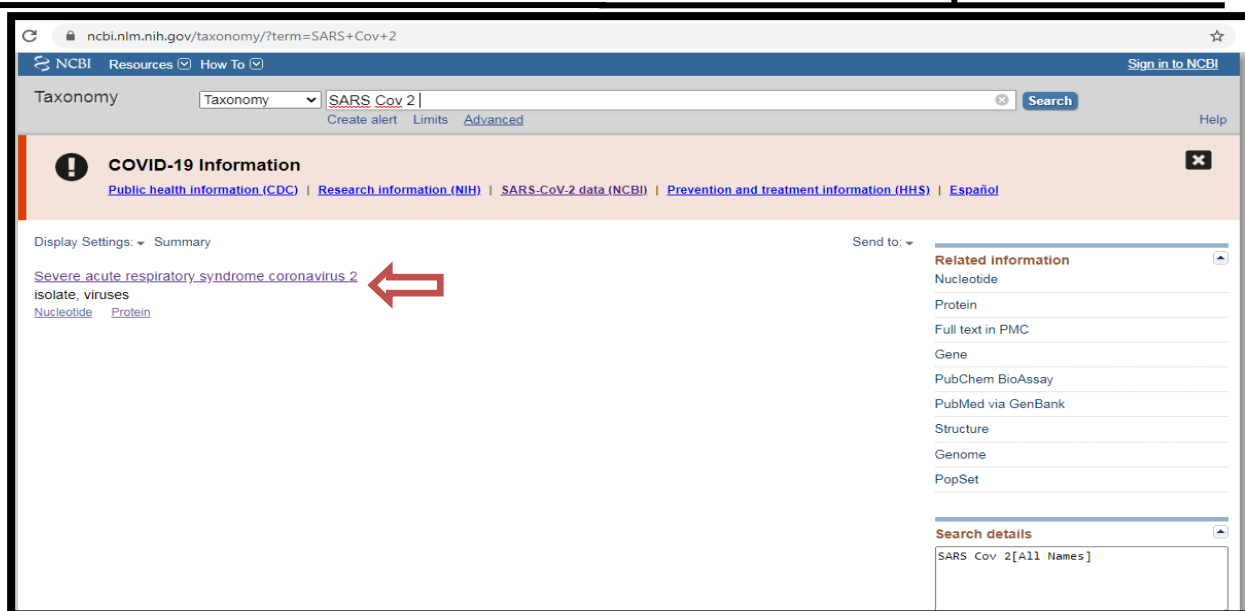


Figure 11 : Analyse de genre « SARS Cov 2» par la GenBank.

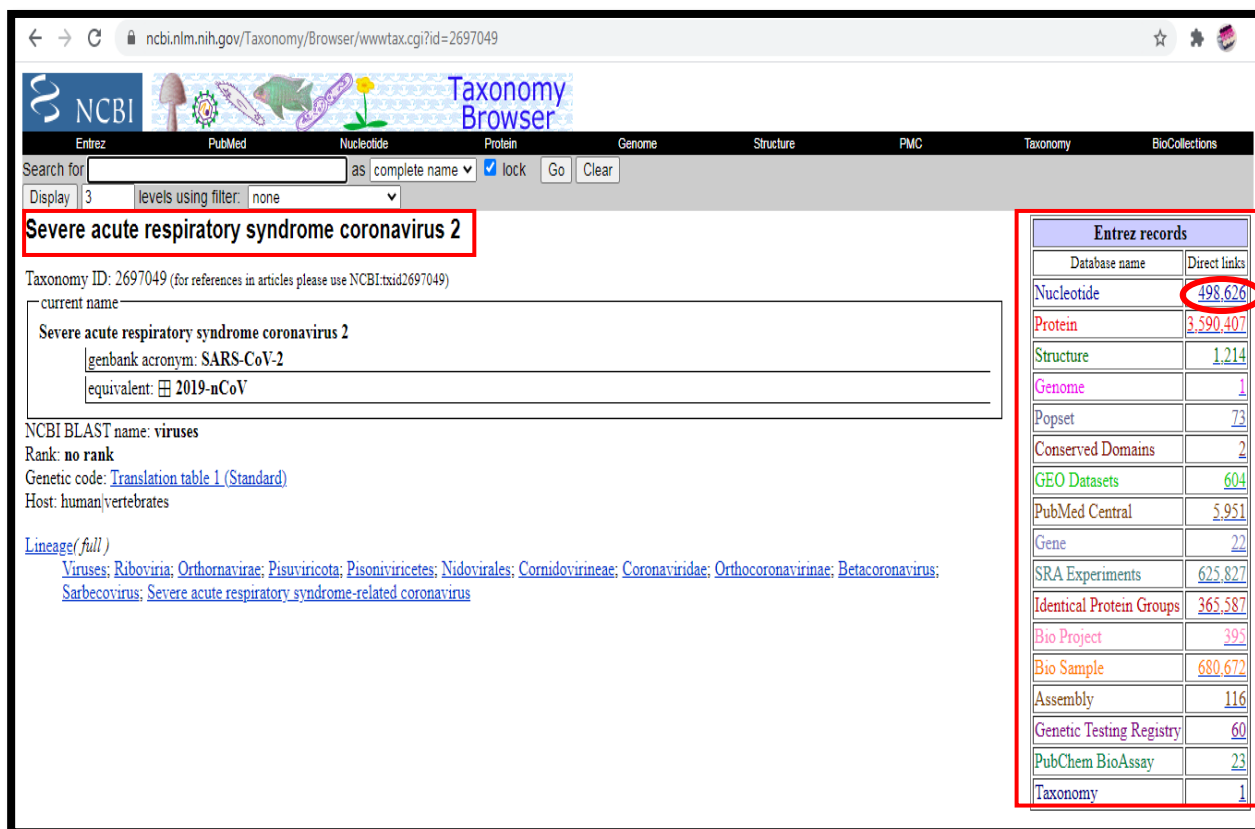


Figure 12 : Tableau informatif de SARS Cov 2 obtenue par la GenBank.

Figure 13 : Exemple des résultats nucléotidiques de SARS Cov 2 obtenue par la GenBank.

5. Nous avons eu toutes les informations sur notre génome (laboratoire, pays, journal, les auteurs ... et tous les gènes étudiés) pour chaque résultat.
6. Les séquences de chaque résultat ont été téléchargées sous forme FASTA à partir de la "GenBank" (Figure 14).

Figure 14 : Exemple d'une séquence nucléotidique sous format FASTA dans la GenBank.

II.3. Traitement des séquences de la GenBank

Il existe plusieurs programmes informatiques pour traiter les séquences, certains sont téléchargés et d'autres en ligne. Les logiciels qui sont utilisés dans notre travail sont : **FaBox**, **MEGA**, **DnaSP**, **DnBA**, **DAMBE**, **NetWork**.

II.3.1. FaBox

FaBox est un ensemble de services Web simples et intuitifs, il permet de faciliter la conversion des séquences vers la forme FASTA.

La conversion des séquences vers la forme FASTA passe par plusieurs étapes :

1. Construire l'URL : <https://users-birc.au.dk/palle/php/fabox/> pour obtenir la page principale de FABOX (Figure 15).

Welcome to FaBox (1.5) - an online fasta sequence toolbox

[\[FAQ\]](#) [FaBox](#)

NEWS:
 21.03.2013 Version 1.5 - Haplotypcollapper working again.
 21.03.2013 I hate my users - now I was forced to code php again..
 19.02.2013 Please stop using fabox - it is outdated!
 19.02.2013 Moved to new server - a few code changes to still work.
 19.02.2013 Added fasta sequence subextractor
 06.08.2011 Small bug in random sequence generator fixed
 29.08.2011 Update with medcili parser
 19.09.2009 FaBox has moved to <http://www.birc.au.dk/software/fabox>
 16.01.2008 Added the FaBox2All_converter.

[Sequence 2 fasta converter \(external tools\)](#)

[FASTA Sequence Conversion Interface](#) - [readSeq at EBI](#)

Working with fasta headers

[Fasta header extractor \(and header splitter\)](#) Simple and fast way of extraction the headers from fasta files - and optionally split each header into fields based on a chosen character word.
[Fasta header editor](#) Simple and fast way of extracting headers, edit them and reapplying them without worrying about the sequence itself.
[Fasta header replacer](#) Some programs do not like the fancy headers in fasta files and you have to live with short, unique names - that are really non-descriptive. Here you can replace headers back and forth by submitting old and new headers - which you'll typically keep in a excel spreadsheet.

Working with fasta datasets/alignments

[Fasta sequence extractor](#) Simple and fast way of extracting some sequences from a large sequence set, based on a list of headers or fuzzy matching.
[Fasta sequence subextractor](#) Simple and fast way of removing some sequences from a large sequence set, based on a list of headers or fuzzy matching.
[Fasta sequence joiner](#) Simple and fast way of joining a set of fasta sequences into one sequence.
[Fasta dataset splitter/divider](#) Simple and fast way of dividing your dataset into two sets by a header keyword. It will split into sets WITH and WITHOUT the given header keyword (like 'females/males', 'population1/population2').
[Fasta alignment joiner](#) Simple and fast way of joining two alignments, sequence by sequence. It will join alignment 1, sequence 1 with alignment 2, sequence 1 and so on... (see example)
[Alignment trimmer](#) Trims an alignment to the shortest sequence. It simply removes the boundary areas that are full of gaps.
[Show variable sites only](#) Extracts all the variable sites from an alignment.
[DNA to haplotype collapser and converter](#) Will collapse a set of sequences into haplotypes and output som simple statistics and formatted input files for Arlequin.

Data conversion

[Strip branch lengths and bootstrap values from network trees \(network parser\)](#) Will take a network tree and create three versions: the original, branch lengths only and topology only (for ASR in panl).
[Create formatted sequence file for PAML analysis \(fasta2paml\)](#) Will format your fasta sequences and create a correct input file for PAML (it's a phylog format with some modifications).
[Fasta2ncel converter](#) Will explode a sequence set into tabular format. It's possible to explode sequences base-by-base and to transpose (flip) the resulting table.
[Create TCS input file from fasta \(fasta2tcs\)](#) Will format your fasta sequences and create a correct input file for the TCS software (TCS: Phylogenetic network estimation using statistical parsimony; Clement et al. 2000).
[Create migrate input file from fasta \(fasta2migrate\)](#) Will format your DNA sequences and create a migrate file called 'infile'. The produced file will be correctly formatted to be analysed by migrate.
[Create MrBayes input file from fasta \(fasta2mbayes\)](#) Will format your DNA sequences and create a MrBayes file, with data and analysis settings included. The produced file will run a series of analyses and produce consensus trees with posterior probabilities.
[DNA to haplotype collapser and converter](#) Will collapse a set of sequences into haplotypes and output som simple statistics and formatted input files for Arlequin.

Sequence generation

[Random DNA sequence generator](#) Generates a specified number of random DNA sequences of given length and exact base composition (will also generate sequences of varying length)

[Afficher le Bureau](#)

Figure 15 : La page principale du programme FaBok.

2. Télécharger le fichier puis cliquer sur (**submit**) pour le convertir (Figure 16).

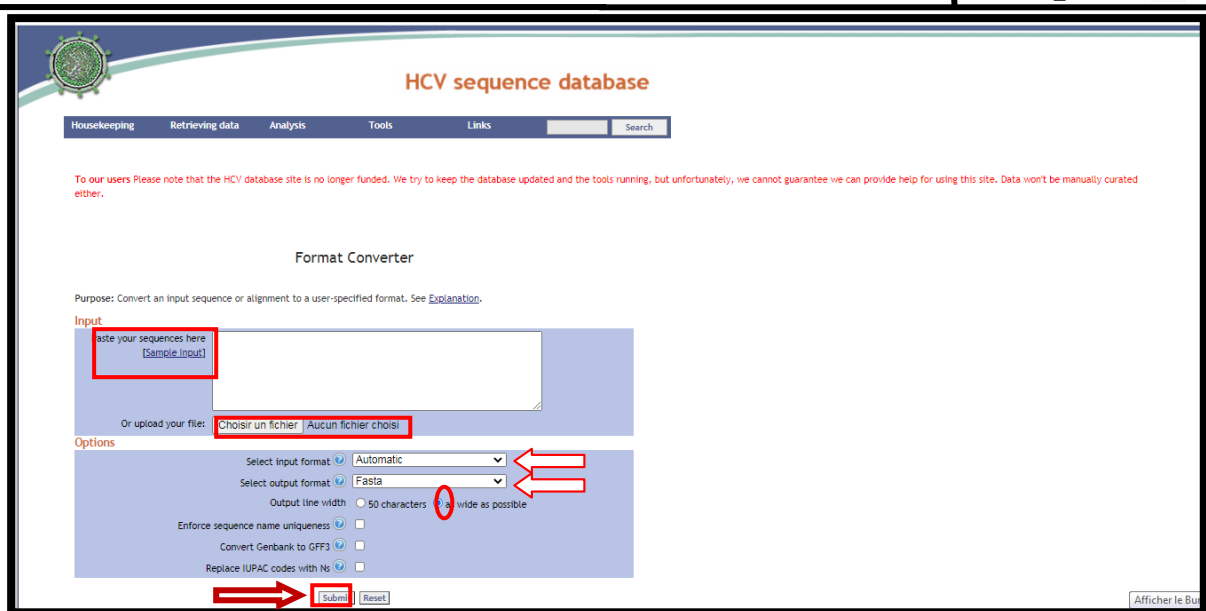


Figure 16 : Conversion des séquences en forme FASTA par l'utilisation de FaBox.

II.3.2. Le programme Adn Fasta format-MEGA_v1.0

Au milieu de notre travail et avec le traitement des milliers de séquences, nous avons constaté que FaBox prend beaucoup de temps, donc nous avons pensé à inventer un nouveau programme pour accélérer notre travail. Avec l'aide d'un informaticien nous avons créé le programme **Adn Fasta format-MEGA_v1.0** qui marche selon les étapes suivantes :

1. L'utilisateur télécharge toutes les séquences des gènes sous forme FASTA à partir de la **GenBank**.
2. Puis, Il crée un nouveau document texte dans le bureau renommé (**adn**) et il transfère toutes les formes FASTA dans ce document.
3. Après avoir cliqué 2 fois sur le programme, ce dernier va automatiquement supprimer toutes les interlignes et les espaces entre les bases et crée un nouveau document texte nommé (**r_adn**) sur le bureau qui contient tous les résultats.

II.3.3. MEGA

MEGA (Molecular Evolutionary Genetics Analysis) est utilisé pour procéder à l'analyse phylogénétique, et pour obtenir à la fin les arbres phylogénétique, cela après le passage par différentes étapes :

La figure 17 représente la barre d'outils de MEGA 11 (La dernière version).

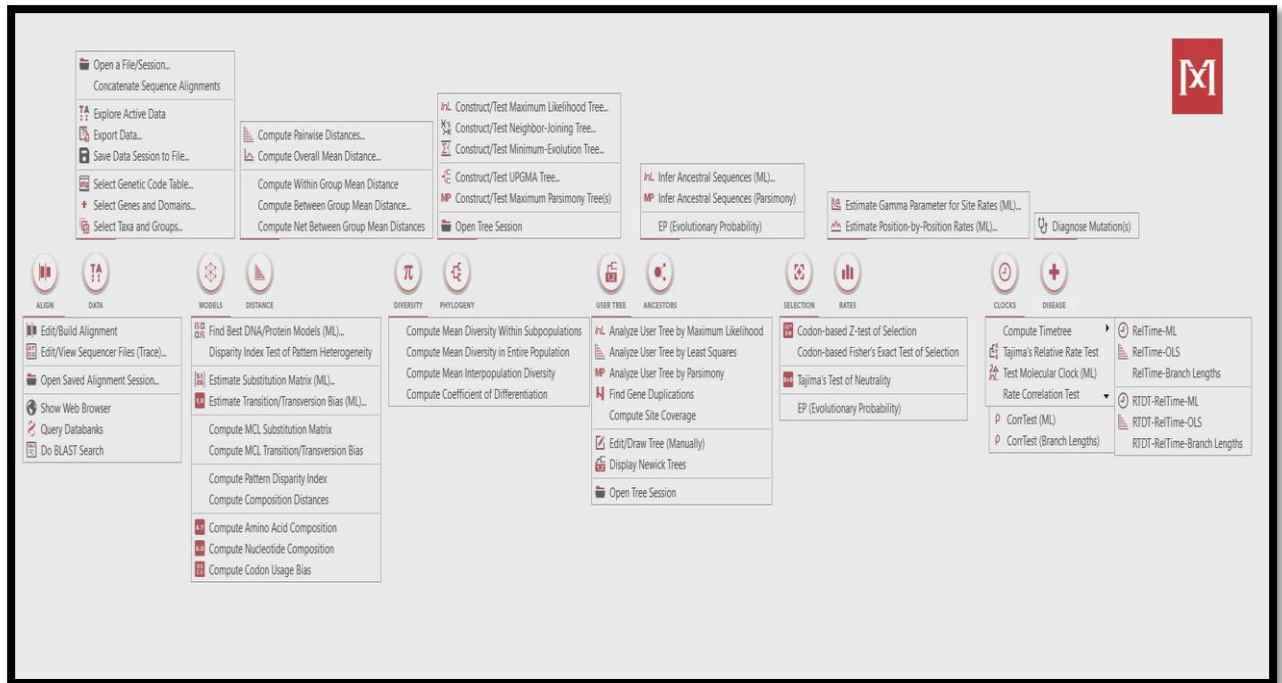


Figure 17 : Barre d'outils de MEGA 11.

1. Insérer les séquences nucléotidiques (Figure 18).

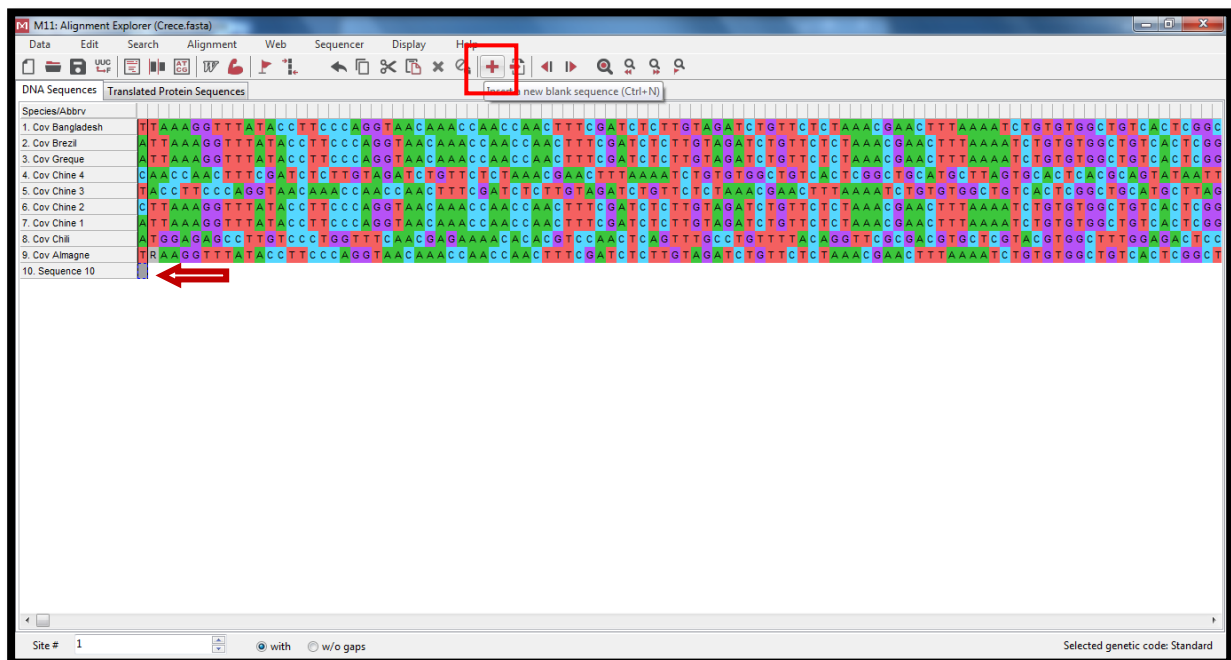


Figure 18 : L'ajout des séquences nucléotidiques sur logiciel MEGA 11.

2. Cliquer sur (W) puis sur (Align DNA) (Figure 19) pour faire l'alignement des séquences par Clustal W.



Figure 19 : L'alignement multiple des séquences nucléotidiques par le programme ClustalW.

3. Appuyez sur (**Phylogeny**) puis (**Construct/Test UPGMA Tree...**) (Figure 20) afin de choisir la méthode de construction de l'arbre phylogénétique.

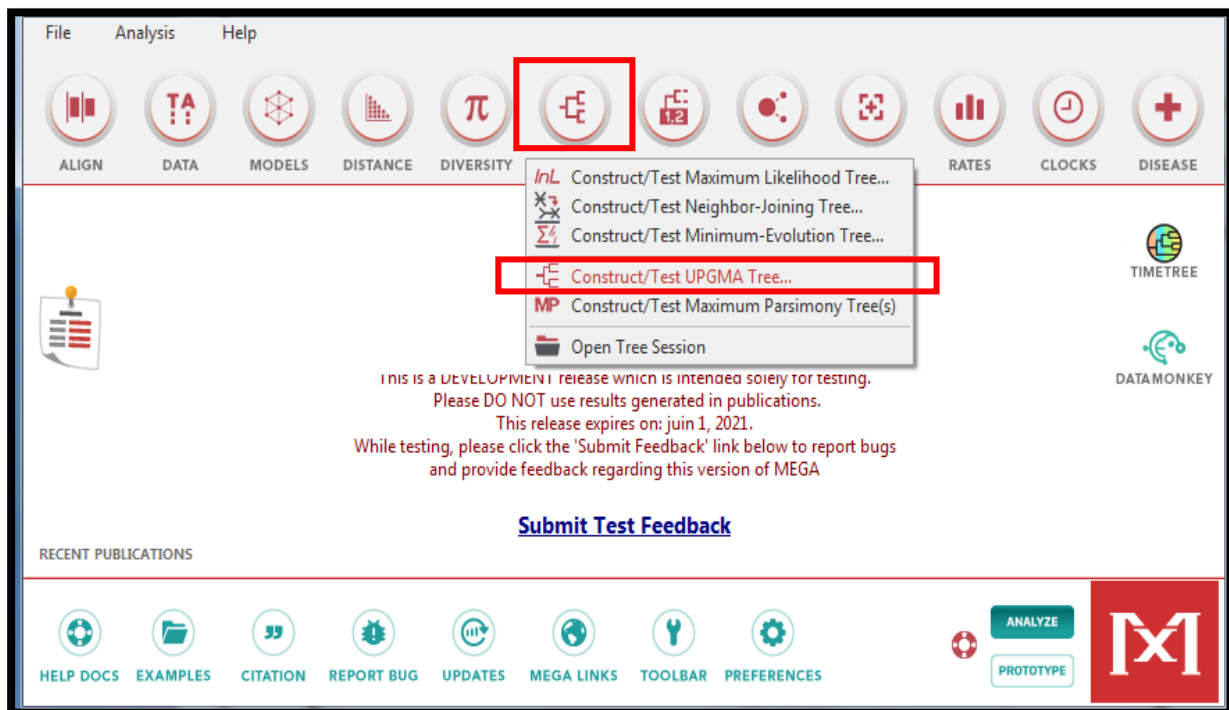


Figure 20 : La méthode UPGMA utilisée pour la construction de l'arbre phylogénétique.

4. L'ajustement des paramètres comme indiqué sur la figure 21. Dans notre cas nous avons utilisé la méthode UPGMA, basée sur les similarités entre paires de séquences (Yoann, 2012) avec un Bootstrap de 1000.

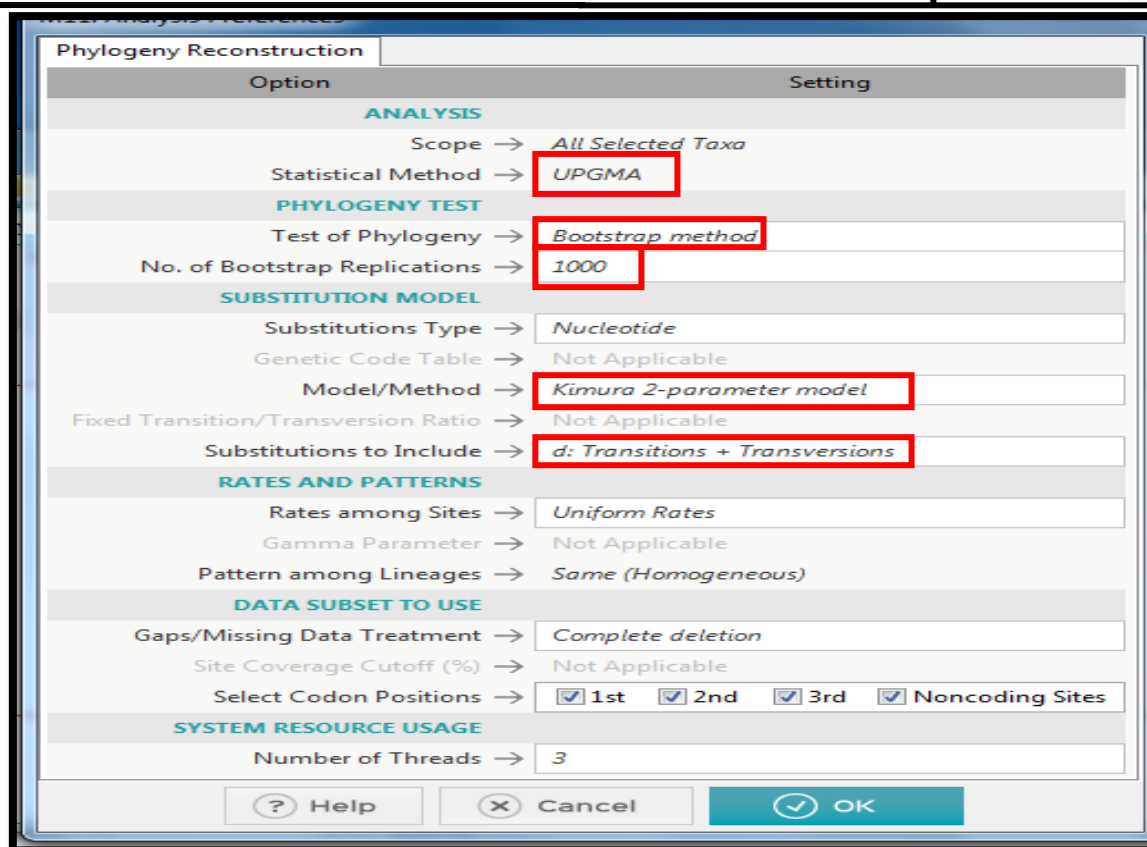


Figure 21 : Les paramètres utilisés pour la reconstruction phylogénétique.

II.3.4. DnaSP

Une fois l'alignement des séquences terminé, nous passons à la détermination et la vérification de la composition nucléotidique par les haplotypes et cela à l'aide de logiciel DnaSP v6.12.03 (Figure 22).

DNA Séquence Polymorphism Analysis of Large Data Sets (DnaSP) est un outil de bioinformatique conçu pour l'analyse complète de variation des données de séquence d'ADN. Le programme permet la caractérisation détaillée des niveaux et des modèles de séquence d'ADN. La version 6 incorpore de nouvelles capacités particulièrement adapté, DnaSP 6 implémente :

- 1) Des modules pour lire et analyser des données à partir de méthodes de partitionnement génomique.
- 2) Des méthodes plus rapides évolutives pour les données de séquençage à haut débit.
- 3) Des statistiques récapitulatives pour l'analyse de données sur la génétique des populations multi-locus (ROZAS *et al.*, 2021).

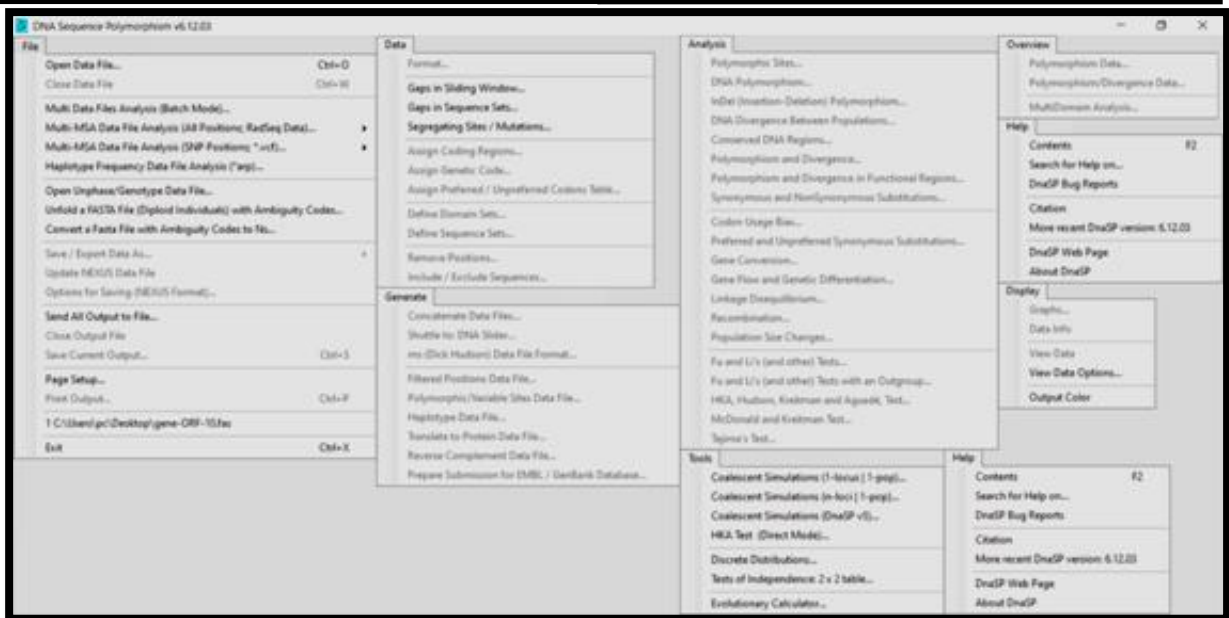


Figure 22 : Barre d'outils de DnaSP v6.

Dans le travail avec logiciel de DnaSP nous avons suivi les étapes :

1. Ouvrir un nouveau fichier DATA pour insérer une séquence nucléotidique sous forme FASTA.
2. Cliquer sur (**Generate**) puis sur (**haplotype data Fil...**) comme indiqué sur la figure 23.
3. Enregistrer les résultats sur des fichiers **Nexus** ou de données de **Roehl**.

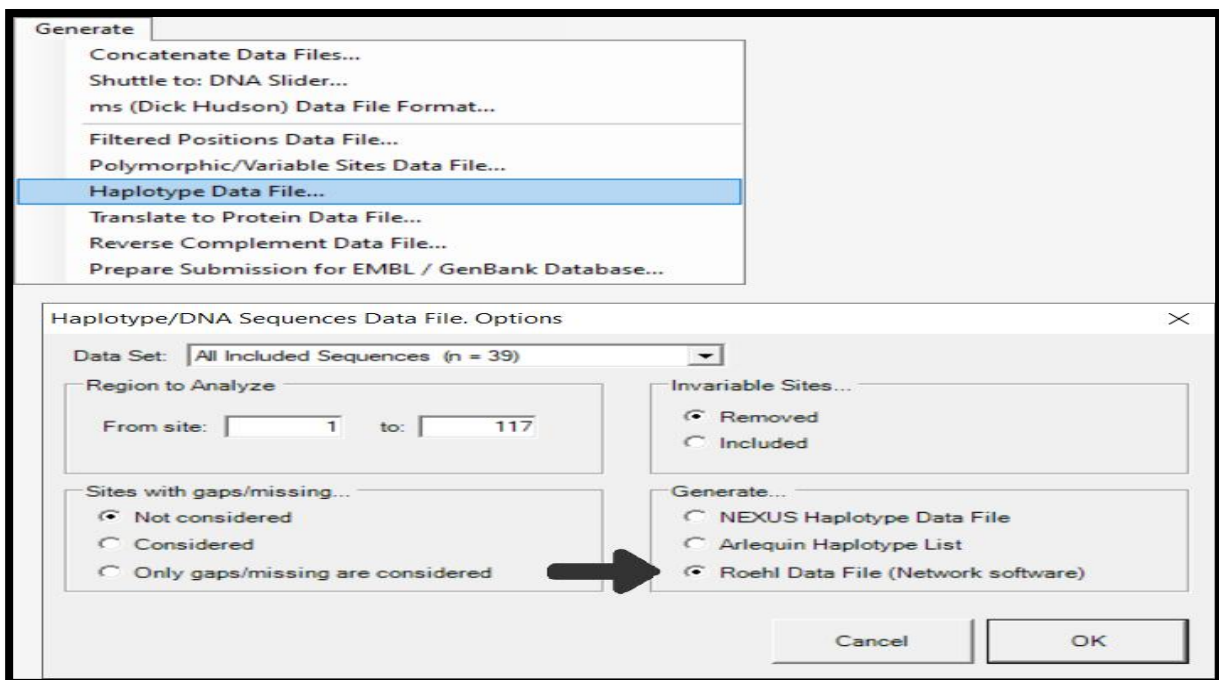


Figure 23 : Interface utilisateur de DnaSP v6.

II.3.5. DNA Basis Analysis (DnBA)

A l'aide d'un informaticien, nous avons créé le programme **DnBA (DNA Basis Analysis)** qui renforce et facilite notre travail et pour ajouter plus d'informations annexés au logiciel **DAMBE** par le calcul des pourcentages des bases azotiques (A, T, C, G) et $(G-C)/(G+C)$, $(A-T)/(A+T)$ et donnant les intervalles des pourcentages des bases, ainsi que le **min** et le **max** de $(G-C)/(G+C)$ et $(A-T)/(A+T)$. Les étapes pour travailler avec ce programme sont :

1. Ouvrir un fichier texte dans le bureau contenant les séquences d'ADN nommé **(a)**.
2. Cliquer 2 fois sur le programme **DnBA**. Les résultats seront trouvés dans le bureau sous forme d'un document texte nommé **(r)**.

II.3.6. DAMBE

DAMBE (Data Analysis in Molecular Biology and Evolution) (Figure 24) est un programme d'informatique utilisé pour l'analyse descriptive et comparative des données moléculaires. C'est un logiciel intégré pour la conversion, la manipulation, la description statistique et graphique, et pour l'analyse des données de séquences nucléotidiques (Evans et al., 2001).

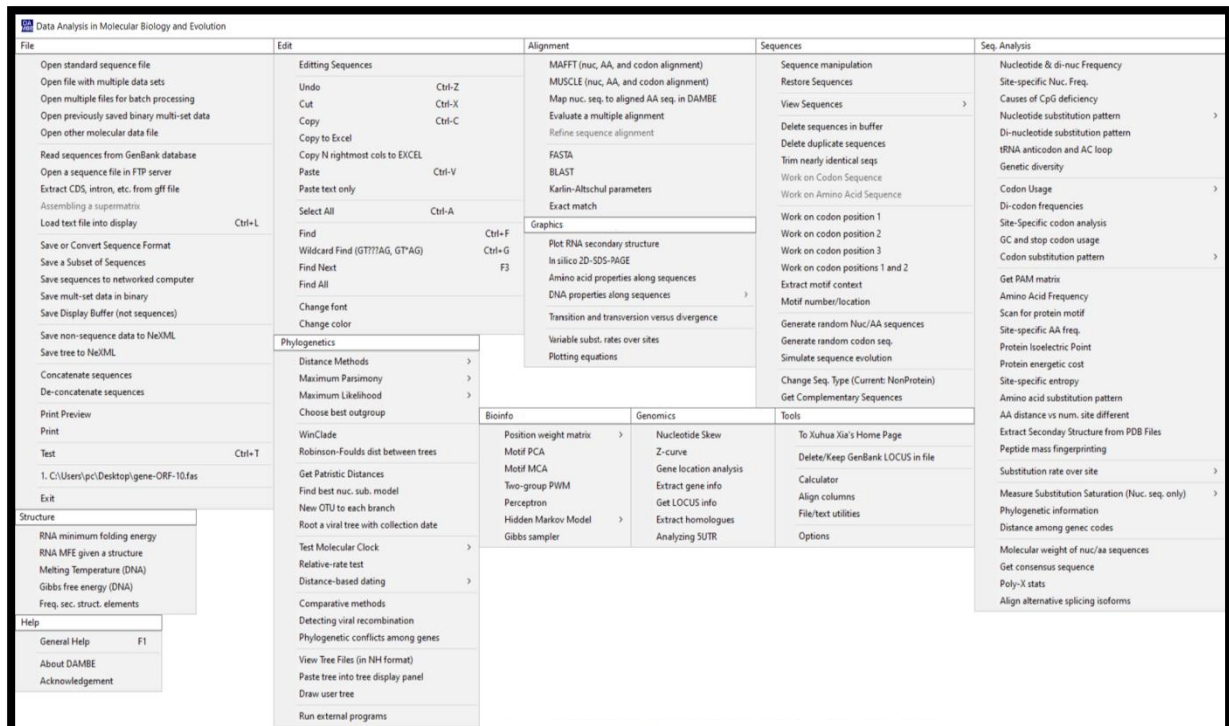


Figure 24 : Barre d'outils du programme DAMBE v7.3.2.

L'utilisation de ce logiciel se fait selon les étapes suivantes:

1. Insérer une séquence nucléotidique (sous forme FASTA) cliquer sur (**File**) puis (**Open standard séquence file**).
2. Concernant le tableau d'informations de la séquence, choisir (**Non-protein Nuc.Seq**) puis cliquer sur (**Go**).
3. A la fin cliquer sur **Genomics** ==>> **Nucleotide Skew** ==>>**Go** Pour obtenir les résultats sous forme de graphes et de statistiques (Figure 25).

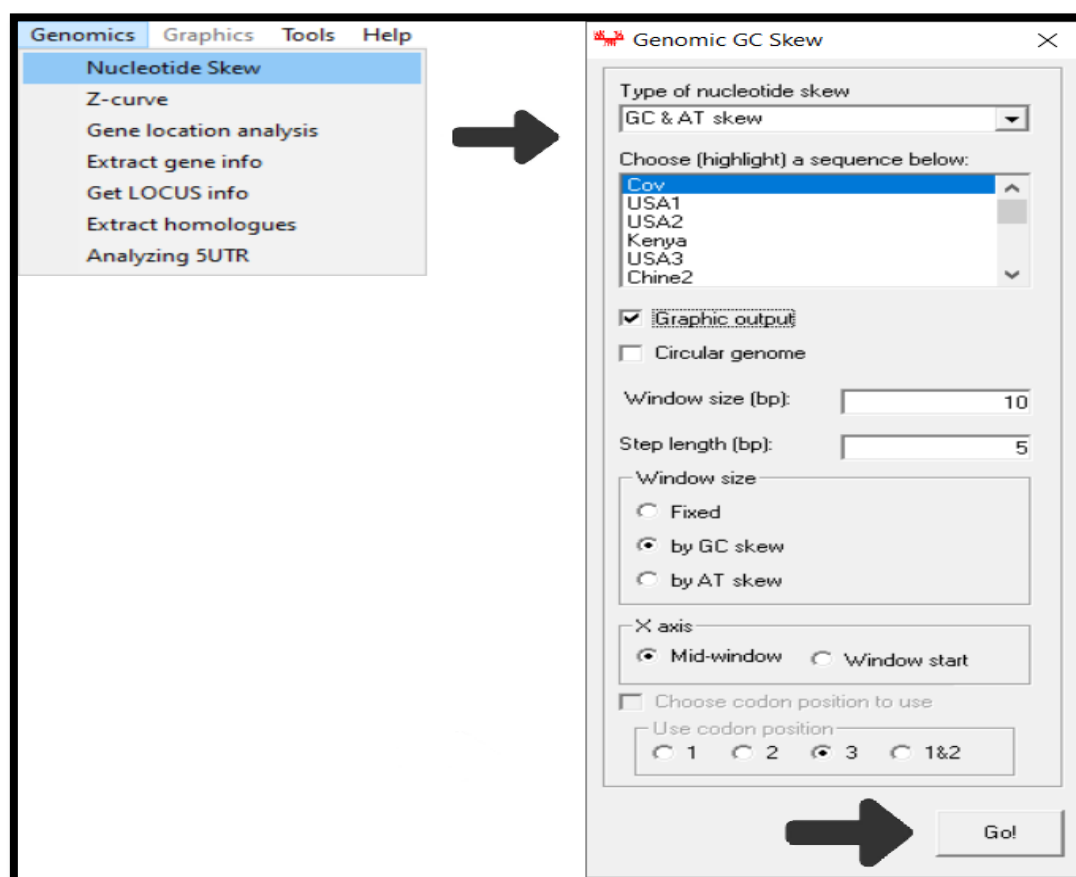


Figure 25 : L'étape d'obtention des résultats statistiques et graphiques de CG /AT.

4. Obtenir des résultats statistiques et graphiques de CG /AT.

II.3.7. NetWork

Le logiciel Network est utilisé pour reconstruire les réseaux phylogénétiques, et déduire les types ancestraux et les types potentiels, les ramifications et les variantes évolutives, et pour estimer les datations. Le réseau génère des arbres et des réseaux évolutifs à partir de données génétiques, linguistiques et autres. Le réseau peut alors fournir des estimations d'âge pour n'importe quel ancêtre de l'arbre (**Huson et Bryant., 2012**).

Nous avons travaillé avec le logiciel NetWork Version 10.2.0.0 (Figure 26), une version très récente utilisée dans les recherches sur SARS Cov2 ; par exemple permettant aux utilisateurs de créer plus facilement des fichiers rdf SARS Cov 2 (**Forster et Forster, 2020**).

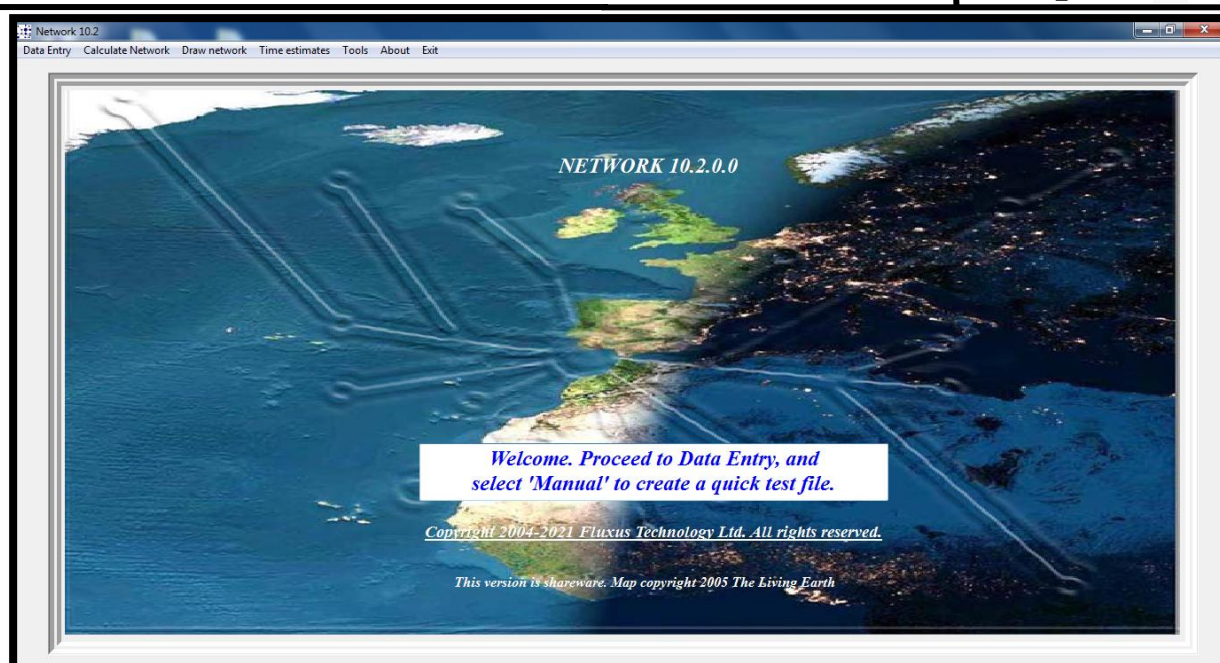


Figure 26 : La page d'accueil de Network v10.2.0.0.

Pour obtenir les différents réseaux de NetWork, nous avons appliqué 2 méthodes, une se base sur les haplotypes et l'autre se base sur les différentes séquences d'ADN directement.

II.3.7.1. Méthode 1 base sur les haplotypes :

1. Afin d'ouvrir un fichier de données **rdf** pour les calculs de MJ NetWork nous avons suivi les étapes présentés dans la figure 27.

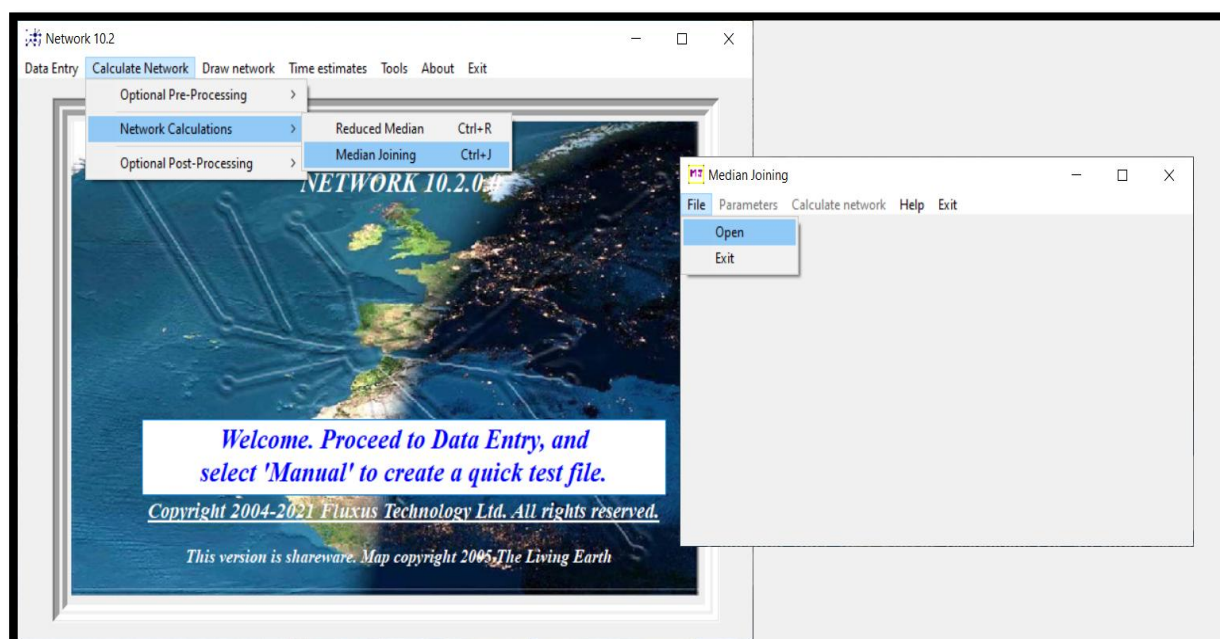


Figure 27 : Ouverture du fichier de données de rdf pour le calcul de MJ NetWork.

2. Cliquer sur (**Calculate network**) puis enregistrer le fichier sous forme **out** (Figure 28).

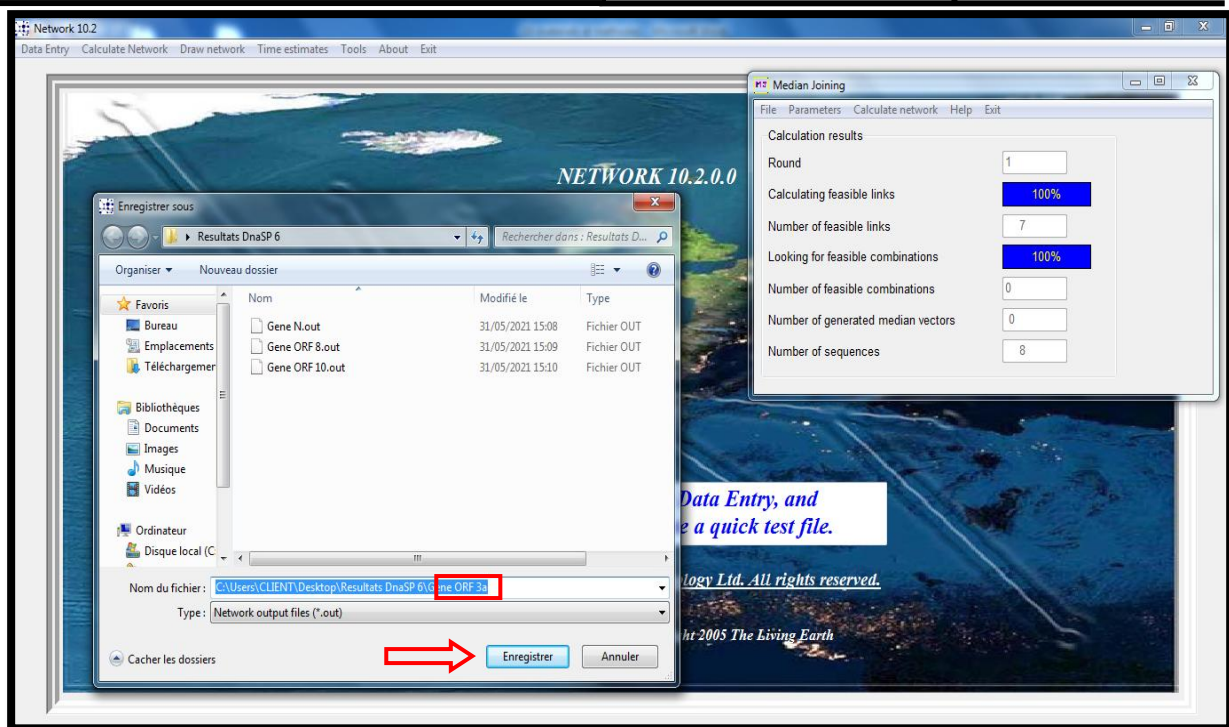


Figure 28 : Enregistrement d'un fichier de calcul MJ sous forme out.

3. Cliquer sur **Draw network** et entrer le fichier précédemment enregistré, pour obtenir le réseau haplotypique, puis l'enregistrer sous forme (**fdi**) (Figure 29).

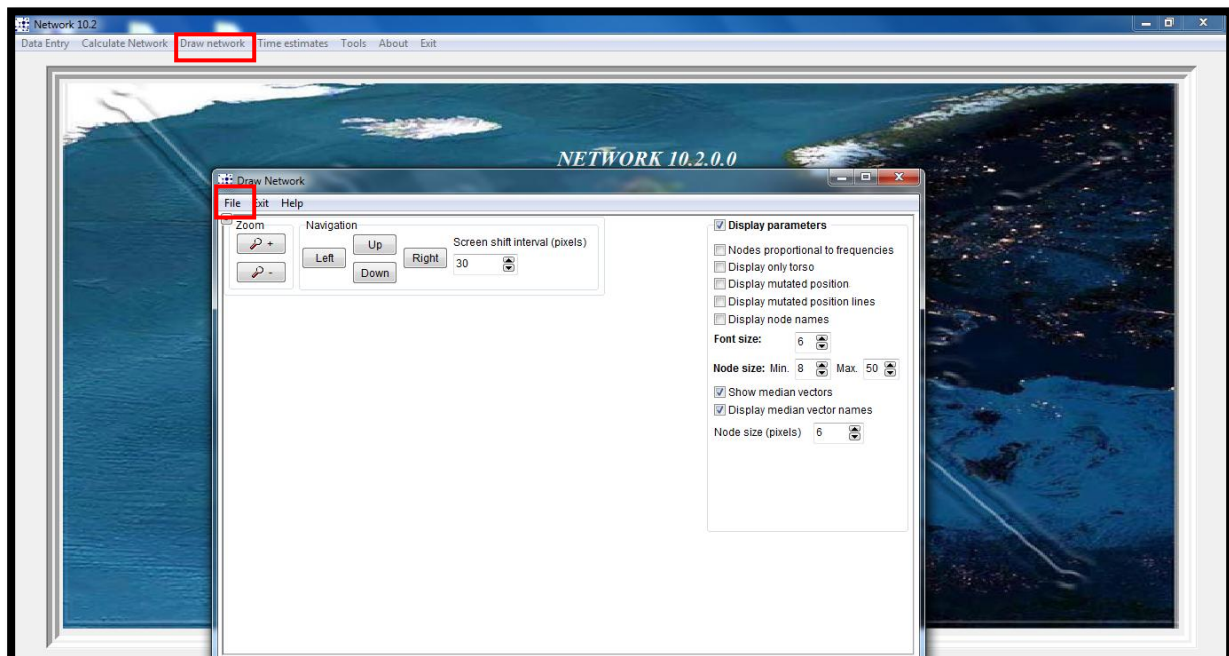


Figure 29 : Les dernières étapes de reconstruction du réseau haplotypique.

II.3.7.2. Méthode 2 basée sur les séquences directement :

Nous avons utilisé le programme **FAS to RDF Conversion** pour transformer les fichiers de forme FASTA en forme **RDF**

1. Transformer les fichiers **MEGA** en forme **FASTA**
2. Construire l'URL suivante : <http://fas2rdf.herokuapp.com/>
3. Insérer en ligne le fichier **FASTA** dans le programme **FAS to RDF Conversion** présenté dans la figure 30 puis cliquer sur **Convert to RDF**

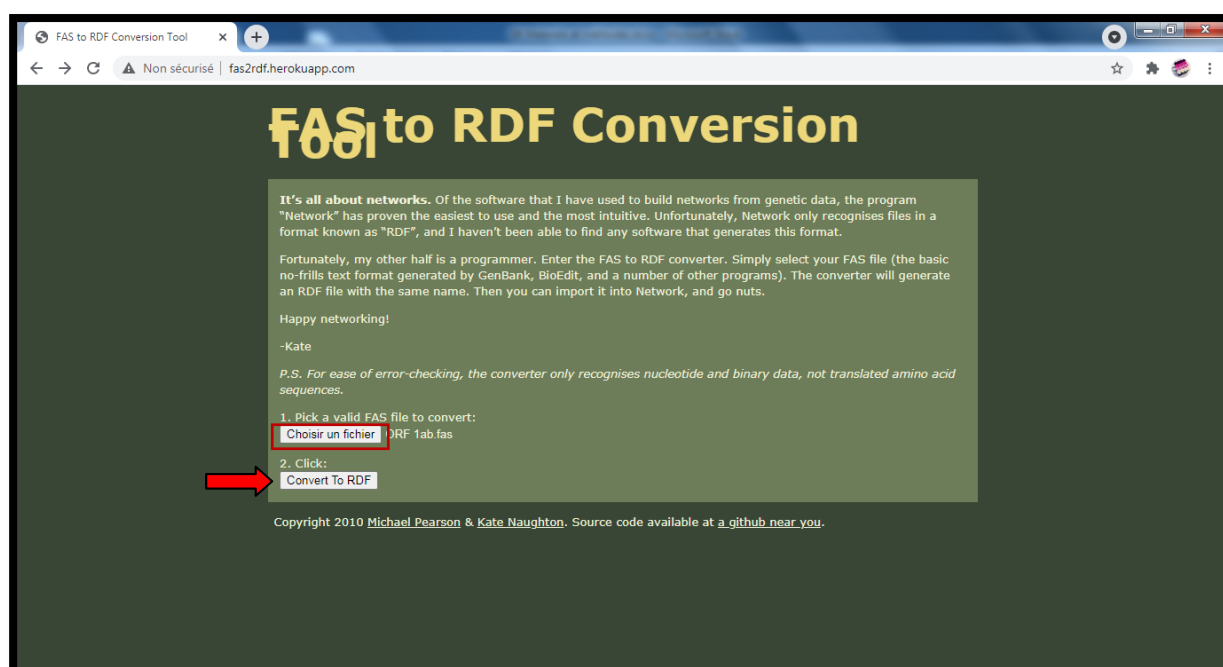


Figure 30 : Page d'accueil en ligne de programme **FAS to RDF Conversion**.

3. Après le téléchargement du fichier **rdf** par ce programme, nous avons suivi les étapes de la méthode 1 pour avoir à la fin le réseau.

Les résultats que nous avons obtenu se basent sur la comparaison des séquences d’ADN des gènes de SARS Cov 2 analysés dans différents laboratoires à travers plusieurs pays dans le monde, publiés sur la GenBank ; le traitement des séquences par des logiciels de bioinformatique nous a permis de construire les arbres et les réseaux phylogénétiques ainsi que la détermination les positions des mutations sur les gènes analysés. En même temps avoir la composition nucléotidique de chaque gène grâce à des analyses statistiques.

III.1. Résultats

III.1.1. Genbank

Nous avons pris 12 gènes de la GenBank de génome de SARS Cov2 (Taille de génome ente 29.8kb et 29.9kb) présentés dans le tableau V ; c’est les gènes les plus étudiés par différents laboratoires dans 45 pays au monde qui sont soit des gènes structurales: M (669pb), N (908pb), S (3822pb), E (228pb) ou bien des gènes fonctionnels: ORF1ab (21290pb), ORF3a (828pb), ORF7a (366pb), ORF7b (132pb), ORF6 (186pb), ORF8 (193pb), ORF10 (117pb), ORFa1 (12142pb) (Rozhgar et al., 2020 ; Sekulic et al., 2020).

Tableau V : Quelques exemples des pays qui ont fait des études analytique des gènes de SARS Cov 2.

Gènes	Tailles	Pays
ORF1ab	[21100pb-21400pb]	Chine, USA, Espagne, Israël, Népal, Corée, Brésil, Iran, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Allemagne, Turquie, Serbie, Grèce, Italie, Egypte, Tunisie, Inde, Pays Bas, Suède, Chili, Sri Lanka, Kazakhstan .
S	[3800pb-3900pb]	Chine, USA, Kenya, Espagne, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Italie, Iran ,Suède, Sri Lanka, Kazakhstan, Pays bas, Chili, Egypte, Tunisie, Inde.
ORF3a	[800pb-	Chine, USA, Kenya, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pérou, France, Bangladesh, Tchèque, Allemagne, Serbie, Grèce, Italie, Egypte, Tunisie, Inde, Suède,

	900pb]	Iran, Sri Lanka, Kazakhstan, Pays bas, Chili.
E	[200pb-300pb]	Chine, USA, Kenya, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Suède Italie, Egypte, Tunisie, Inde.
M	[600pb-700pb]	Chine, USA, Kenya, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Iran ,Sri Lanka, Kazakhstan, Pays bas, Chili, Suède, Italie, Egypte, Tunisie, Inde.
ORF6	[150pb-200pb]	Chine, USA, Kenya, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Italie, Egypte, Tunisie, Inde.
ORF7a	[300pb-400pb]	Chine, USA, , Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Italie, Iran ,Sri Lanka, Kazakhstan, Pays Bas, Chili, Suède ,Egypte, Tunisie, Inde.
ORF7b	[100pb-150pb]	Chine, USA, Espagne, Israël, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Iran, Sri Lanka ,Kazakhstan, Pays bas, Chili, Egypte, Tunisie.
ORF8	[350pb-400pb]	Chine, USA, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Iran, Sri Lanka, Kazakhstan, Pays bas, Chili Italie, Egypte, Tunisie, Inde, Suède.
N	[1000pb-1500pb]	Chine, USA, Kenya, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Inde, Italie, Egypte, Tunisie. Iran, Sri Lanka, Kazakhstan, Pays Bas, Chili, Suède.
		Chine, USA, Kenya, Espagne, Israël, Colombie, Corée, Brésil, Taiwan, Vietnam, Finlande, Pakistan, Pérou, France, Japon, Bangladesh, Tchèque, Allemagne, Turquie, Serbie, Grèce, Italie,

ORF10	[90pb-140pb]	Egypte, Tunisie, Inde, Iran, Sri Lanka, Kazakhstan, Pays bas, Chili, Suède.
ORF1a	[13100-13400]	Pakistan, Pérou

III.1.2. Résultats de FaBox

Les séquences que nous avons utilisé dans les différents logiciels devaient être sous la forme FASTA, pour cela nous avons utilisé le FaBox afin de convertir les séquences vers cette forme. La figure suivante (31) représente un exemple de résultat de FaBox.

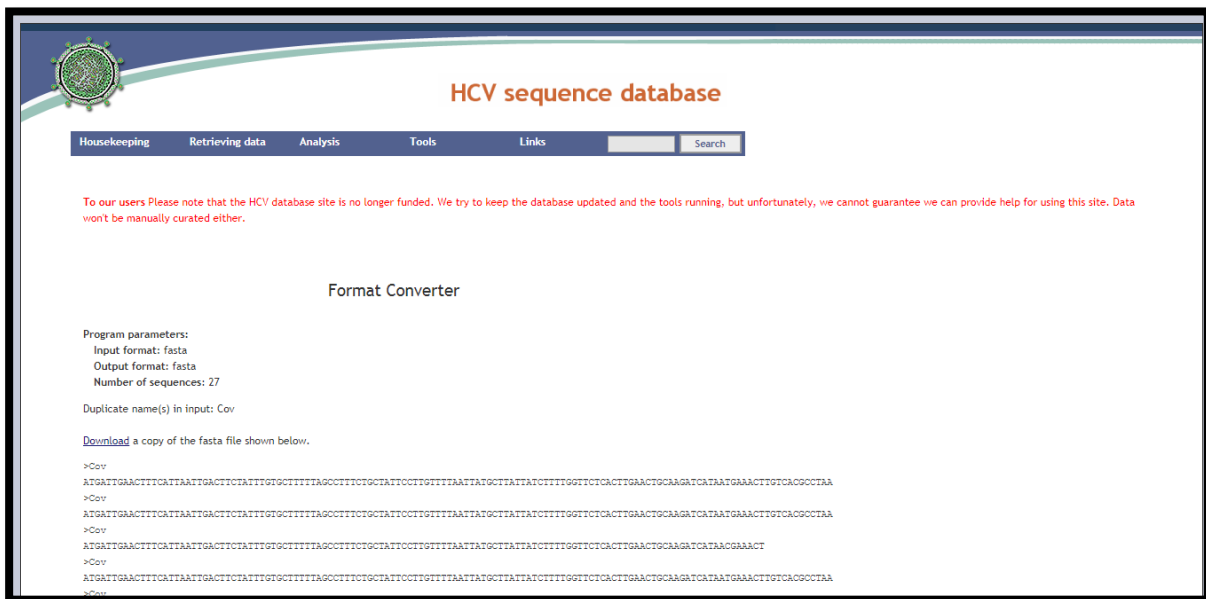


Figure 31 : Un exemple de forme FASTA obtenue par FaBox.

III.1.3. Résultats de programme Adn Fasta format-MEGA v1.0

Ce programme nous donne directement la forme FASTA alignée sans espaces entre les bases et les lignes. La figure 32 montre le résultat que nous a donné ce programme.

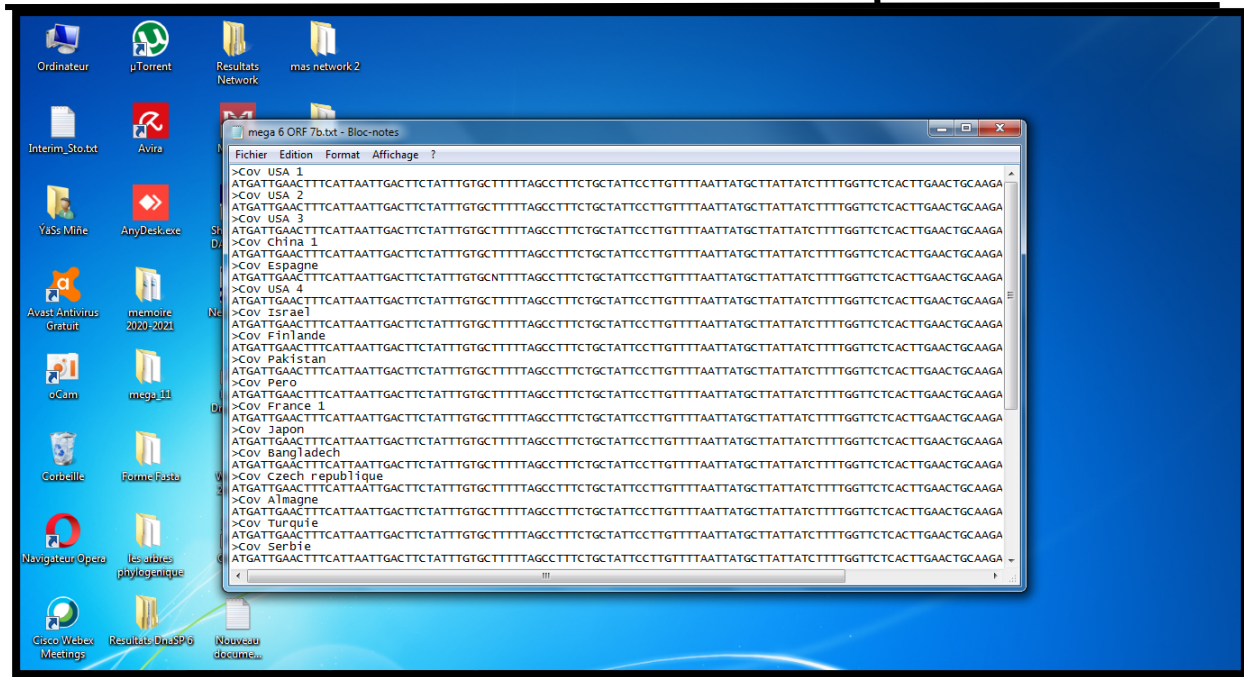


Figure 32 : Résultat du programme Adn Fasta format-MEGA v1.0.

III.1.4. MEGA

Nous avons obtenu différents arbres phylogénétiques par MEGA, le tableau VI résume les résultats des arbres phylogénétiques des gènes de SARS Cov2.

Tableau VI : Résumé des résultats des arbres phylogénétiques obtenus par MEGA des gènes de SARS Cov2.

Racine	Branches	Nœuds	Feuilles
ORF1ab	30	3	10
S	15	3	25
ORF3a	8	3	36
E	43	0	0
M	43	0	0
ORF6	41	0	0
ORF7a	39	0	0
ORF7b	27	2	26
ORF8	3	2	25
N	16	3	30
ORF10	3	1	37

Après l'analyse de tous les arbres phylogénétiques nous avons remarqué qu'il y a deux types d'arbres. 1^{er} type plus compliqué et riche en ramifications (exemple présenté dans la figure 33) ; 2^{ème} type pauvre en ramifications (exemple présenté dans la figure 34). La totalité des arbres son représentés dans l'annexe.

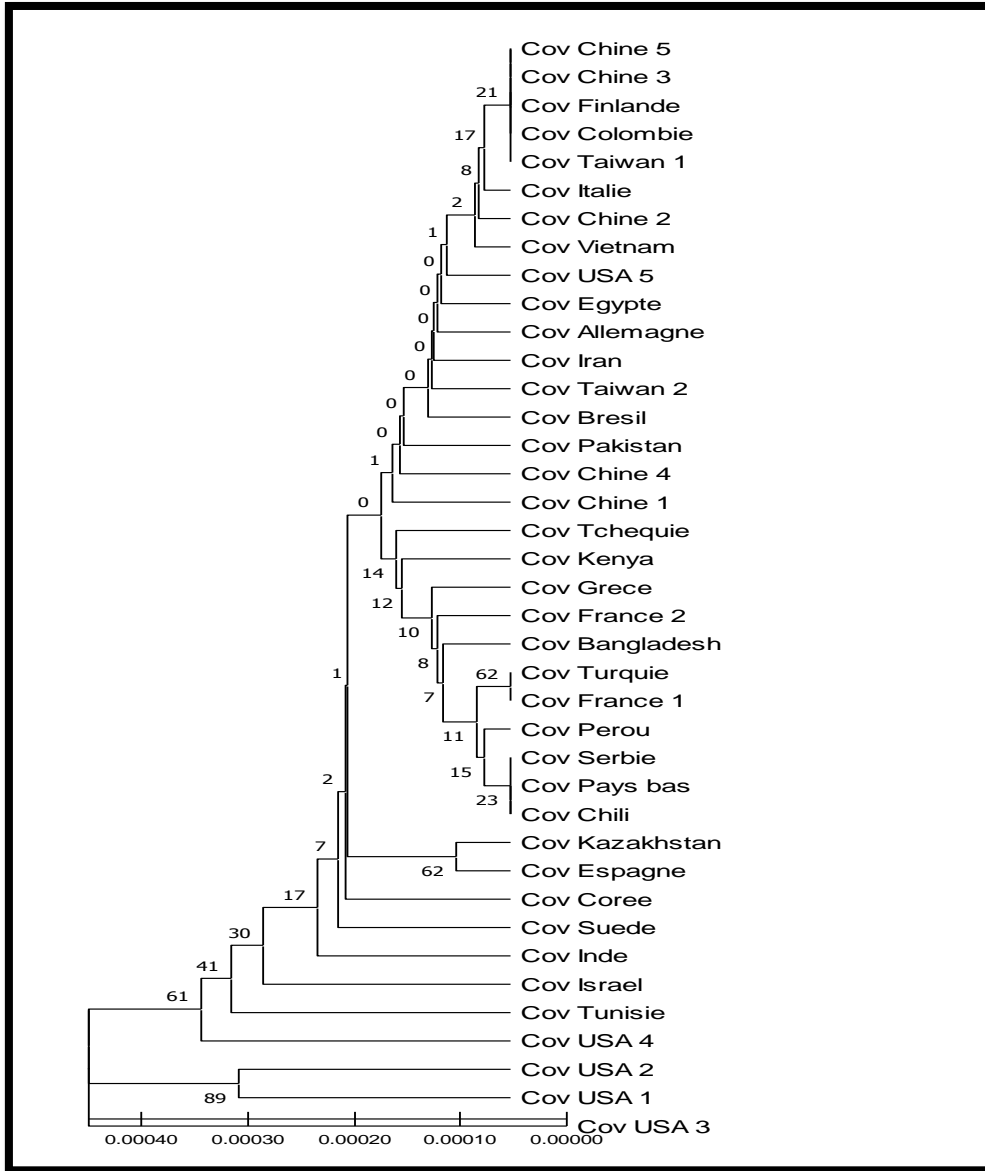


Figure 33 : Arbre phylogénétique de gène ORF1ab de SARS Cov 2 construit à partir de 37 séquences, en utilisant la méthode UPGMA par logiciel MEGA.

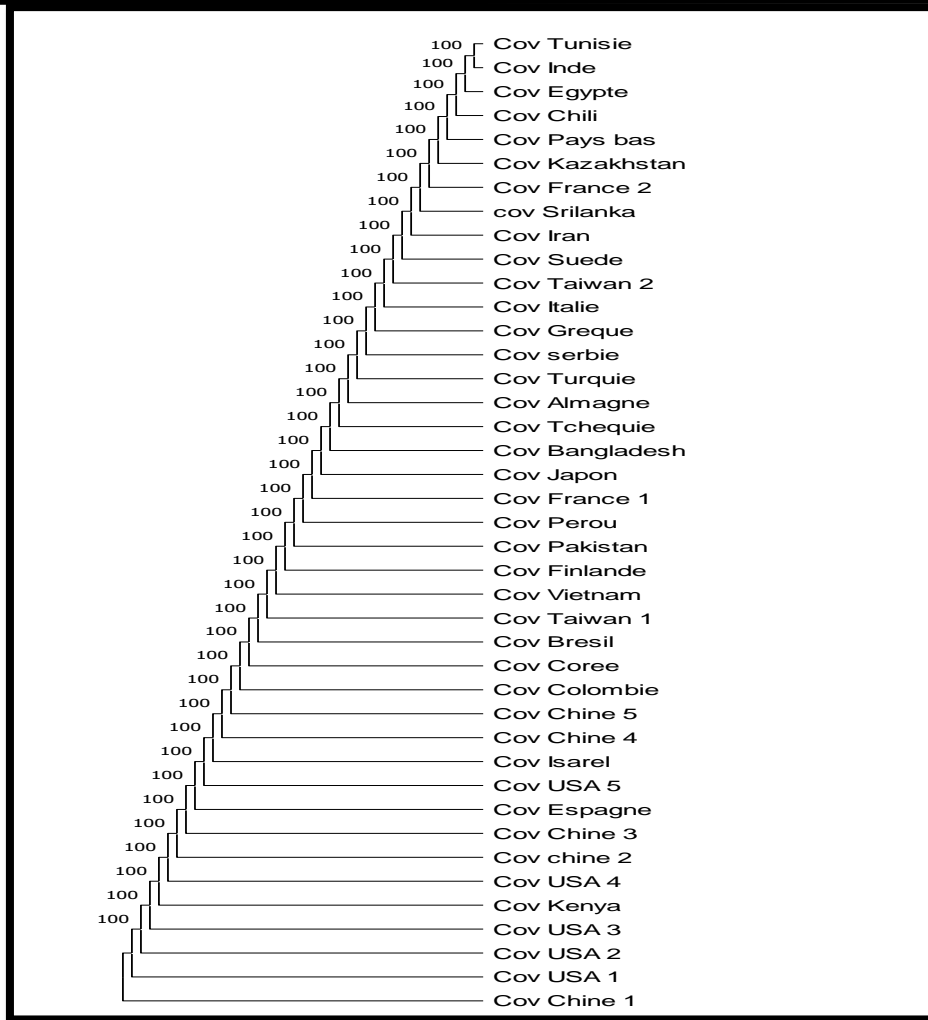


Figure 34 : Arbre phylogénétique de gène ORF 6 de SARS Cov 2 construit à partir de 41 séquences, en utilisant la méthode UPGMA par logiciel MEGA.

III.1.5.DnaSP

Les résultats de DnaSP des gènes étudiés de SARS Cov2, sont organisés soit par des groupes sous forme des haplotypes avec une diversité haplotypique significative positive qui sont les gènes : ORF1ab, ORF3a, ORFa1, ORF10, N, ORF8, S, ORF7a, ORF7b représentés dans tableau VII (Exemple la figure 35 : le gène ORF1ab) ou sous forme d'un haplotype unique ($h=1$) et une diversité haplotypique égale à zéro ($hd=0.000$) et cela dans les cas des gènes E, M et ORF6 (Exemple la figure 36 : le gène E).

Tableau VII : Les résultats de DnaSP des gènes de SARS Cov2.

Gènes	Séquences utilisées	Nombre total de sites	Nombre de sites variables	Nombre d'haplotypes (h)	Diversité haplotypique (Hd)
ORF1ab	37	21281	146	30	0.9790
ORF3a	41	21255	98	29	0.9778

ORFa1	2	13218	7	2	1.000
ORF10	39	117	2	3	0.1012
N	43	1260	23	16	0.7431
ORF8	26	120	1	2	0,0769
S	37	3748	24	15	0.7867
ORF7b	26	120	1	2	0,0769
ORF7a	39	366	1	2	0,0513

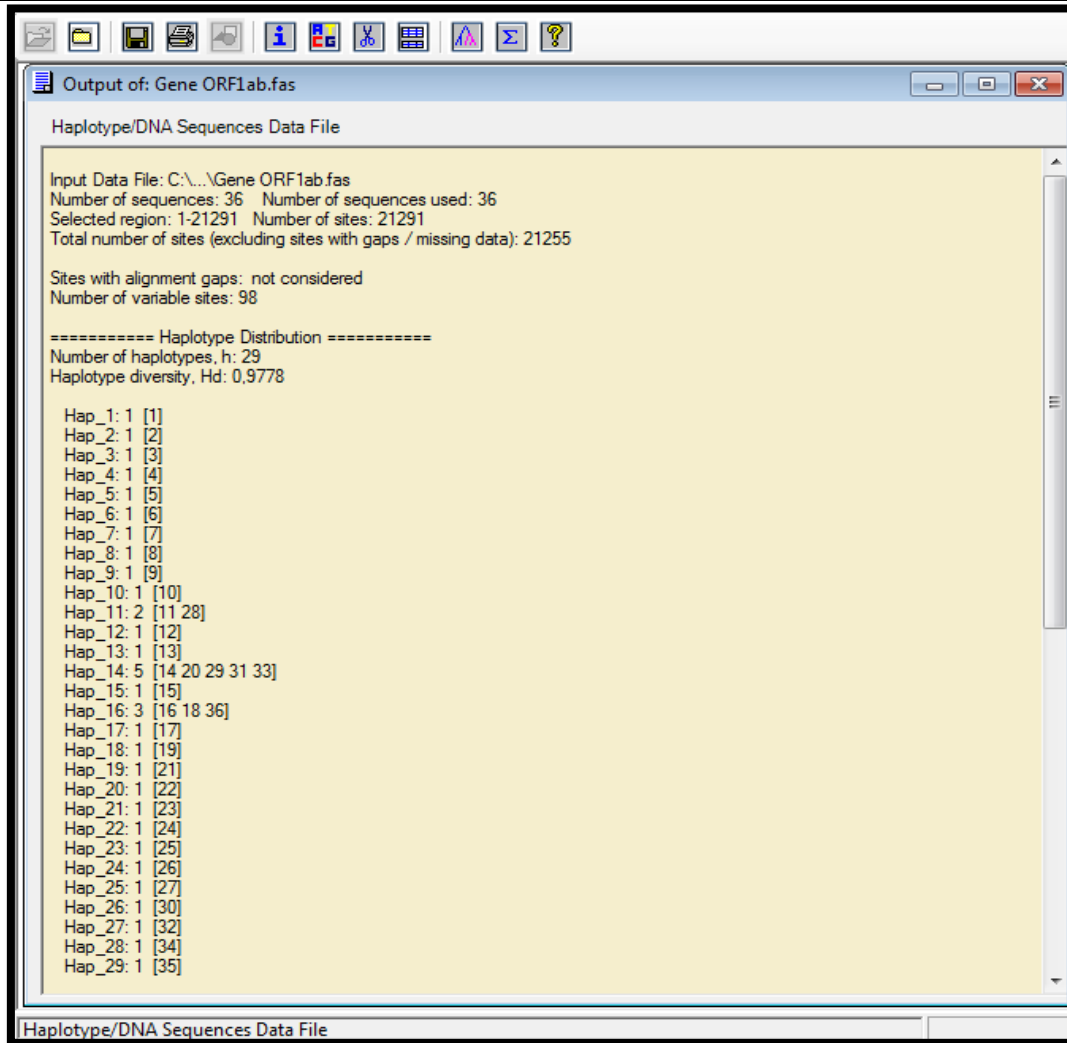


Figure 35 : Nombre des haplotypes de gène ORF1ab calculé par le logiciel DnaSP.

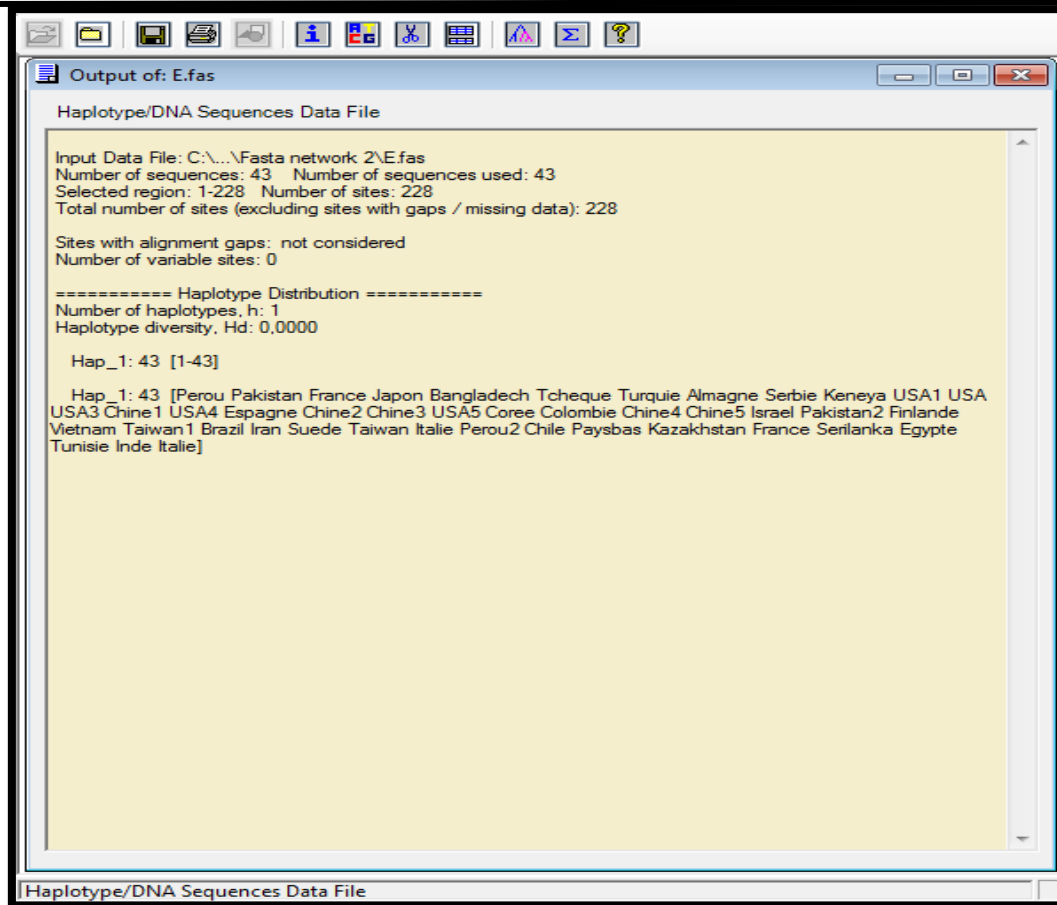


Figure 36 : Nombre des haplotypes de gène E calculé par le logiciel DnaSP (haplotype unique).

III.1.6.DnBA (DNA Bases Analyses)

Nous avons utilisé ce programme afin d'accélérer notre travail par le calcul des pourcentages de : chaque base azotique (A, C, G, T), de $(C-G/C+G)$, et de $(A-T/A+T)$ ceci pour chaque séquence étudiée ; en parallèle il calcul l'intervalle général de chacune des bases (A, T, C, G), et les Min et Max valeurs de $(C-G/C+G)$, $(A-T/A+T)$ pour chaque gène étudié de SARS Cov 2. Nous avons obtenu deux types de résultats :

- Les séquences des gènes E, M, ORF6, avec des pourcentages de : (A, T, C, G), $(C-G/C+G)$ et de $(A-T/A+T)$ identique pour toutes les séquences étudiées du même gène obtenu de différents pays, par exemple le résultat représenté dans la figure 37 pour le gène E.
- Les séquences des gènes S, N, ORF1ab, ORF3a, ORF7a, ORF7b, ORF10, ORF8, ORF1a avec de différents pourcentages de : (A, T, C, G), $(C-G/C+G)$ et de $(A-T/A+T)$ pour toutes les séquences étudiées du même gène obtenu de différents pays ; la figure 38 un exemple de ce genre de résultat (gène S).

Concernant les résultats des autres gènes nous les avons présentés dans l'annexe.

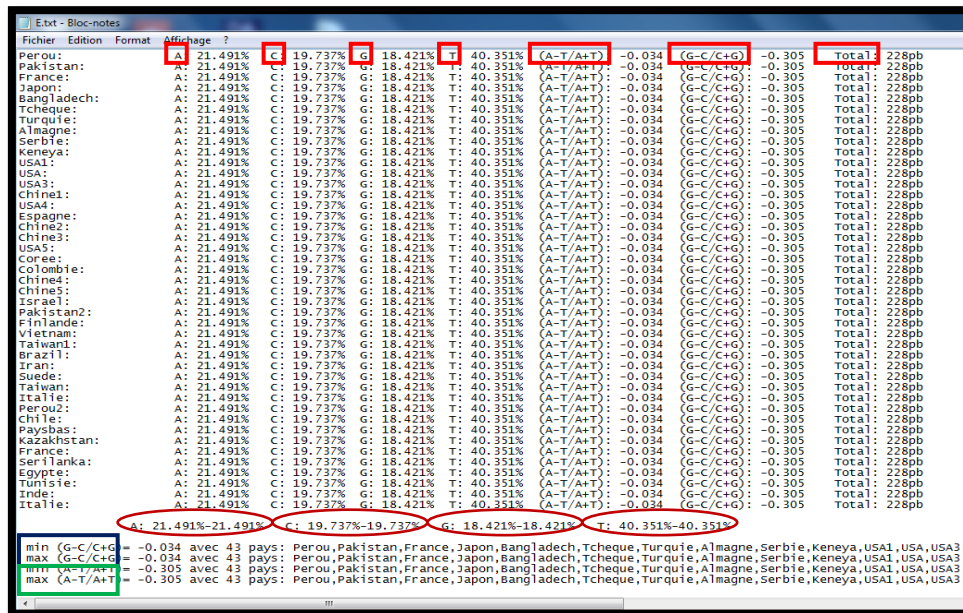


Figure 37 : Analyse des bases azotiques des séquences de gène E par DnBA.

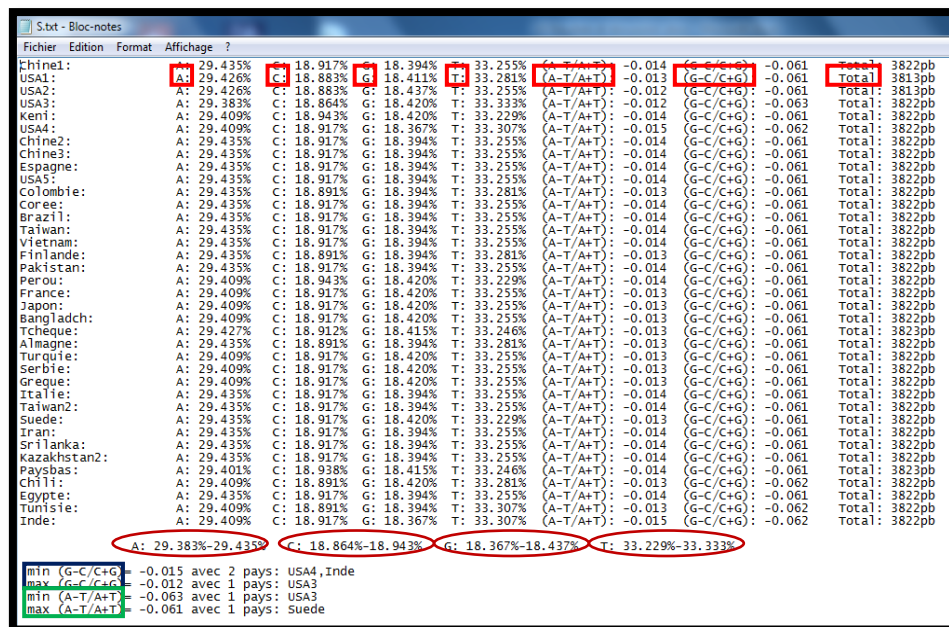


Figure 38: Analyse des bases azotiques des séquences de gène S par DnBA.

III.1.7. DAMBE

Sachant que DAMBE est un logiciel spécialisé dans les calculs et les statistiques moléculaire et en se basant sur les résultats de logiciel précédant (DnBA), qui nous a permis de choisir les gènes les plus variables qui sont : S, N, ORF1ab et ORF3a ; nous

avons appliqué le logiciel DAMBE sur ces quatre gènes ; les résultats de ce logiciel sont présentés sous 3 formes qui sont :

1. Les séquences identiques de chaque gène. La figure 39 est un exemple du gène ORF1ab.
2. Calculs pour chaque gène (S, N, ORF1ab, ORF3a) : $(G-C)/(G+C)$, $(A-T)/(A+T)$, AT skews (la somme $(A-T)/(A+T)$) et CG skews (la somme, $(G-C)/(G+C)$) ; le tableau VIII présente le CG et le AT skews pour les quatre gènes étudiés. La figure 40 montre un exemple des tableaux donnés par DAMBE.
3. La forme graphique de CG et AT skews pour les gènes S, N, ORF3a, ORF1ab ; un exemple de graphe de CG et de AT skews de séquence de Vietnam de gène ORF1ab présenté dans la figure 41.

Tableau VIII : Les variations globales AT et CG Skew des gènes S, N, ORF3a et ORF1ab de SARS Cov 2.

Calculs Genes	Global CG Skews		Global AT Skews	
	Moins variable (Min)	Plus variable (Max)	Moins variable (Min)	Plus variable (Max)
ORF1ab	Bresil	USA1	USA3	Espagne
AT/ CG Skews	0,06083	0,06226	-0,03573	-0,03469
S	USA4	USA3	USA3	Suede
AT/ CG Skews	-0,01474	-0,01193	-0,06255	-0,06010
ORF3a	Coree	Srilanka	Tcheque	Srilanka
AT/ CG Skews	-0,06748	-0,06098	-0,10359	-0,10000
N	USA3	Turquie	USA2	Espagne
AT/ CG Skews	-0,06937	-0,05565	0,19940	0,20540

```

The following sequence pairs are identical to each other:
Serbie..Paysbas
Taiwan..Colombie
Turquie..France1
Taiwan..Finlande
Colombie..Finlande
Taiwan..Chine5
Colombie..Chine5
Finlande..Chine5
Taiwan..Chine3
Colombie..Chine3
Finlande..Chine3
Chine5..Chine3
Serbie..Chili
Paysbas..Chili
    
```

Data Analysis in Molecular Biology and Evolution

File Edit Alignment Sequences Seq. Analysis Phylogenetics

Nucleotide skews for sequence: Bresil
 Global GC Skew = 0.06083
 Global AT Skew = -0.03507
 Sliding window size optimized with GC skew.
 Window size = 1086; stepsize = 106
 Circular genome: No

Window	Mid-window	(G-C)/(G+C)	(A-T)/(A+T)
0	543	0.0757	-0.0117
1	649	0.0698	-0.0016
2	755	0.0814	-0.0280
3	861	0.0713	-0.0080
4	967	0.0738	-0.0203
5	1073	0.0826	-0.0092
6	1179	0.0816	-0.0167
7	1285	0.0733	-0.0407
8	1391	0.0993	-0.0444
9	1497	0.1209	-0.0549
10	1603	0.1064	-0.0618
11	1709	0.1132	-0.0363
12	1815	0.1031	-0.0045
13	1921	0.0813	0.0060
14	2027	0.1127	0.0284
15	2133	0.0862	0.0046
16	2239	0.1137	0.0000
17	2345	0.1671	0.0227
18	2451	0.1682	0.0517
19	2557	0.1401	0.0586
20	2663	0.1306	0.0977
21	2769	0.1678	0.0799
22	2875	0.1559	0.0643
23	2981	0.1518	0.0373
24	3087	0.1499	0.0427
25	3193	0.1479	0.0771
26	3299	0.1089	0.0767
27	3405	0.1050	0.0671
28	3511	0.1156	0.0843
29	3617	0.1061	0.0696
30	3723	0.1122	0.0540
31	3829	0.0707	0.0522
32	3935	0.0606	0.0348
33	4041	0.0354	0.0493

Figure 39 : Résultats DAMBE, les séquences de paires identiques de gène ORF1ab de SARS Cov 2.

Figure 40 : Les analyses statistiques d'une séquence de gène d'ORF1ab par logiciel DAMBE.

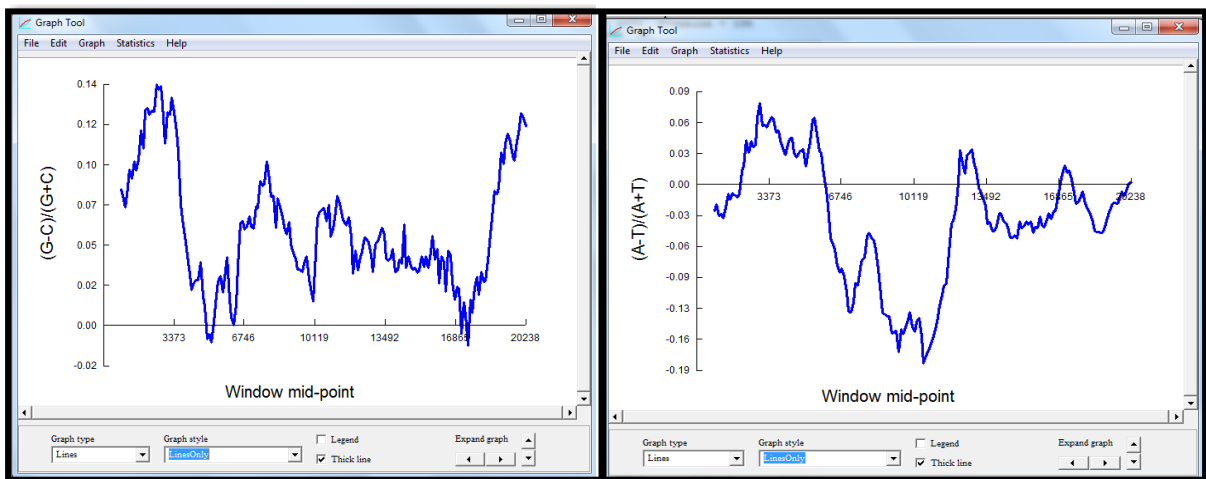


Figure 41 : La variation de GC et AT skews de gène ORF1ab (séquence de Vietnam) obtenu par logiciel DAMBE.

Concernant les résultats de DAMBE pour les séquences des gènes ORF1ab, N, S, ORF3a présentés dans le tableau VIII, sont dans l'annexe figures : 46, 47, 51, 52, 56, 57, 59, 60.

III.1.8. NetWork

Vu l'intérêt des haplotypes qui traduit les différentes séquences nous avons appliqué deux méthodes pour construire les réseaux phylogénétiques des gènes de SARS Cov2.

III.1.8.1. Résultats de la méthode 1:

C'est une méthode basée sur les haplotypes, nous avons utilisé les résultats de DnaSP pour obtenir les réseaux haplo-typiques des gènes de SARS Cov2, les résultats sont:

1. Différents réseaux d'haplotypes pour les gènes N, ORF1ab (figure 42), ORF3a, ORF10, S.
2. Haplotype unique pour les gènes E, M et ORF6 donc pas de réseaux haplotypique.
3. Pas de réseaux pour les gènes ORF7a, ORFa1, ORF7b et ORF8 car ils contiennent moins de 3 séquences différentes.

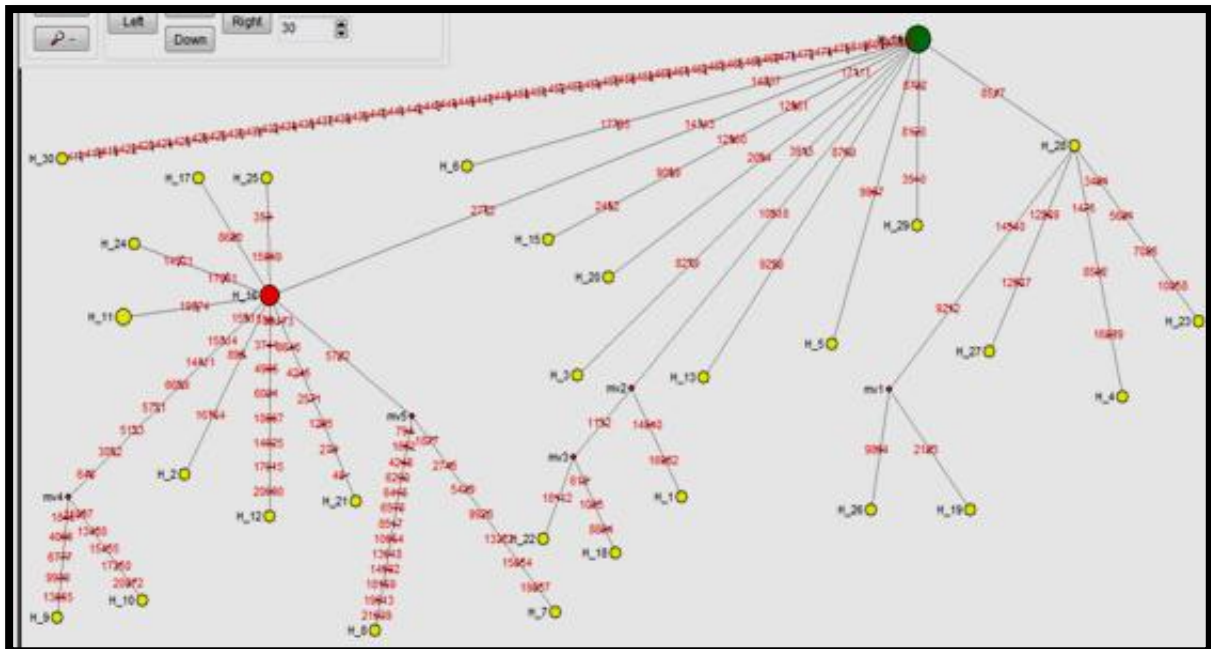


Figure 42: Réseau phylogénétique des séquences de Gene ORF1ab de SARS Cov2, reconstruit par logiciel NetWork à partir de 37 séquence, représentés par des cercles colorés, chaque couleur représente un pays.

III.1.8.2. Résultats de la méthode 2:

Cette méthode se base sur les différentes séquences directement, nous avons utilisé la forme FASTA (résultat de MEGA) qui donne les séquences de différents pays de chaque gène de SARS Cov2. Les résultats obtenus par cette méthode sont :

1. Les gènes E, M, ORF6, ORF7b, ORF8, ORF7a, ORFa1 contiennent moins de 3 séquences différentes donc pas de réseaux.
2. De différents réseaux pour les gènes ORF1ab (Figure 42), N, S, ORF10, ORF3a.

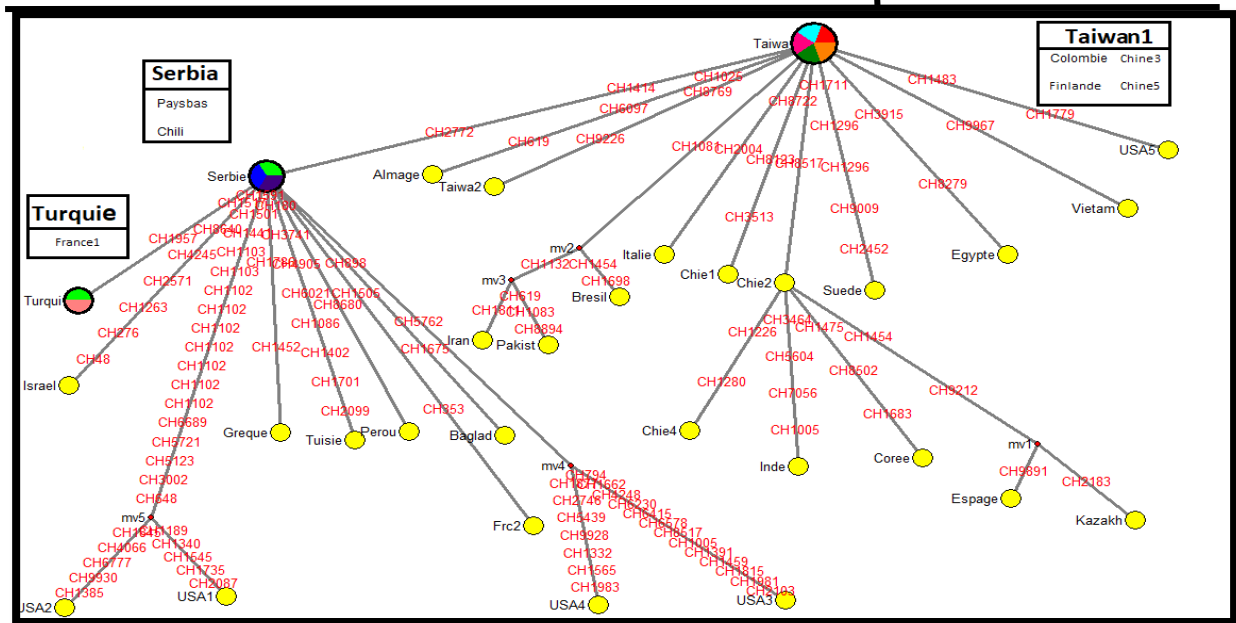


Figure 43: Réseau phylogénétique de séquences de Gene ORF1ab de SARS Cov2, reconstruit par NetWork à partir de 37 séquences, représentées par des cercles colorés, chaque couleur représente un pays.

Nous avons présenté les résultats des autres gènes dans l'annexe.

III.2. Discussion

III.2.1. Comparaison général entre les logiciels utilisés

Après l'analyse des gènes de SARS Cov 2 par les logiciels : MEGA 11, DnaSP 6, DnBA, DAMBE et Network 10 (avec 2 méthodes), nous avons remarqué que les gènes étudiés sont divisés en 3 groupes :

Premièrement les gènes les plus variables qui sont le gène S, N, ORF1ab et ORF3a représenté par des arbres phylogénétiques donnés par la méthode UPGM riche en ramifications (les figures 33, 44, 48, 53), selon **Mercatelli et Giorgi (2020)** ces gènes font partie des gènes les plus susceptibles de mutations, en parallèle les réseaux phylogénétiques de ces gènes construits par logiciel Network représentés dans les figures 43, 45, 49, 54 dans lesquels on voit des cercles soit avec couleurs uniques jaune qui présentent une seule séquence, ou par des cercles avec plusieurs couleurs ce qui traduit l'inclusion de plusieurs séquences, ces cercles sont les nœuds dans les arbres phylogénétiques qui se lient entre eux avec des branches qui représentent les distances génétiques (mutations) entre eux, ce qui est défini dans les travaux de **Rasmont (1997), Gattolliat (2002) et Schmidt (2003)** ou ils ont appelé la relation entre deux nœuds par branche. Les branches peuvent être évaluées, c'est-à-dire que l'on peut leur associer une mesure (une distance, une

quantité d'évolution, un nombre de mutations). Cette distance qui exprime la présence des mutations est traduite par la détection de plusieurs haplotypes pour le même gène ce qui est défini par une diversité haplotypique (H_d) que nous pouvons détecter grâce à plusieurs logiciels entre autre le DnaSp (Anderson et Karel, 2011).

Deuxièmement les gènes les moins variable qui sont le gène ORF8, ORF10, ORF7a, ORF1a et ORF7b avec une diversité haplotypique (H_d) calculé par DnaSp 6 entre 0,0513 et 0.1012 par rapport aux gènes les plus variable qui ont une diversité haplotypique entre 0.7431 et 0.9790 ; Xia (2001) explique que cette diversité haplotypique élevée dû au taux élevé des mutations de type transition ou transversion ; en parallèle Roy et al. (2020) ont trouvé que les gènes de SARS Cov 2 subissent des transitions et transversions ce qui explique leurs diversité haplotypique.

Troisièmement les gènes non variés qui sont le gène M, E et ORF6, toutes les séquences qui représentent chaque gène sont identique, pour cela ils ne présentent pas une diversité haplotypique ($H_d = 0$) car ils ont un haplotype unique exprimé par des arbres phylogénétiques simple avec des distances génétiques identique (Figures 34, 71, 72). Comme les gènes moins variables, les gènes non variables ne présentent pas des réseaux haplotypique Network parce qu'ils contiennent moins de 3 séquences différentes sauf le gène ORF10 qui présente un réseau phylogénétique Network (Figure 62), ce que nous avons remarqué dans son arbre phylogénétique (Figure 61) qui contient 3 branches donc 3 différentes séquences.

Pour le gène ORFa1 nous n'avons trouvé sur la Genbank que 2 séquences, ces séquences ont donné une diversité haplotypique élevée (1.000) car elles sont différentes entre elles ; à cause de ce faible nombre de séquences nous n'avons pas pu construire pour ce gène l'arbre phylogénétique et le réseau Network.

A l'aide du programme DnBA, nous avons comparé les % des bases A, T, C et G pour les différentes séquences obtenues de différents pays, nous avons remarqué que les résultats (Figures 38, 50, 55, 58) pour les gènes S, N, ORF3a et ORF1ab qui subissent des mutations, exposent des pourcentages différents des bases ce qui concorde avec les travaux de Xia (2013), Xia et Xie (2001) qui expliquent que les variations de A, T, C et G signifie qu'il y a des mutations. En même temps DnBA nous a calculé $(A-T/A+T)$ et $(A-G/A+G)$ des séquences de chaque gène et cela nous a facilité le travail avec le logiciel DAMBE afin de construire directement les graphes des séquences les plus et les moins variables.

Sachant que GC skew positif représente la richesse de G sur C et le GC skew négatif représente la richesse de C sur G (Tillier et Collins., 2000), les résultats graphiques de DAMBE pour les valeurs minimales de AT skew (globale $A-T/A+T$) et de CG skew (globale

C-G/C+G) ainsi que les valeurs maximales de AT skew (globale A-T/A+T) et de CG skew (globale C-G/C+G) indique que pour les gènes les plus variables qui sont : le gène S présenté dans les figures 46, 47 est riche en cytosine par rapport à la guanine et en thymine par rapport à l'adénine ; en parallèle les graphes présentés dans les figures 56, 57 pour le gène N, il y a une richesse de la cytosine sur la guanine et de l'adénine sur la thymine. En même temps les figures 51, 52 de gène ORF3a montre aussi qu'il y a une richesse de la cytosine sur la guanine et de la thymine sur l'adénine. Alors que pour les graphes dans les figures 59, 60 de gène ORF1ab, nous avons remarqué qu'il y a une richesse de la guanine sur la cytosine et de l'adénine sur la thymine. Pour les gènes les moins variables ORF6, ORF10, ORF7a et ORF7b représentés dans les figures 63, 65, 67, 70 respectivement nous avons remarqué que les pourcentages de A, T, C, G sont proches entre eux. Dans les figures 37, 73, 74 qui représentent les gènes E, M, ORF6 qui sont les gènes non variés, les pourcentages des bases azotiques A, T, C, G sont identiques pour toutes leurs séquences. **Koyama et al. (2020)** ont prouvé que les variations de C, G, A et T signifient que la séquence a subi des transitions et transversions c.à.d. des mutations chose qui nous a permis dans notre travail sur les gènes qui représentent des variations de déterminer des haplotypes. En parallèle **Graw et al. (2006)** a prouvé que plus l'ADN est riche en paires G/C plus l'ADN 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.

III.2.2. Comparaison entre les séquences de SARS Cov 2 de différents pays

D'après **Wang et al. (2021)** le coronavirus 2019 (COVID-19) a été signalé à Wuhan, dans la province chinoise du Hubei au début de 2020. L'agent pathogène de COVID-19 a été isolé et identifié comme un nouveau coronavirus le 7 janvier 2020, et a ensuite été nommé par le comité international de classification des virus comme le coronavirus 2 du syndrome respiratoire aigu sévère (SARS Cov 2) ; selon **Awadasseid et al. (2021)** la Chine a utilisé le gène S pour confirmer la nature de ce virus qui a causé une pandémie mondiale au cours des 2 dernières années.

Après l'analyse des résultats obtenus, nous avons constaté que les séquences de la Chine dans les arbres phylogénétiques des gènes : S (Figure 44), ORF3a (Figure 48), N (Figure 53),

ORF1ab (Figure 33), ORF8 (Figure 64), ORF7b (Figure 66), ORF7a (Figure), et ORF10 (Figure 61), sont dans les mêmes branches avec d'autres pays, ce qui est expliqué par la similitude entre les séquences, et qu'ils ont subi les mêmes mutations, par exemple nous avons dans l'arbre phylogénétique de gène S, la séquence de la Chine et la séquence de Pakistan sont dans la même branche, même dans le réseau phylogénétique de gène N (Figure 45) se trouvent dans le même cercle, chose qui a été expliquée par **Umair et al. (2021)** ou il a trouvé que la séquence de Pakistan est une séquence de variant B.1.1.7 qui présente une mutation au niveau du gène N. Nous avons aussi la séquence de Pérou pour le gène S se trouvant dans la même branche avec la séquence chinoise, et c'est ce que **Aguilar-Gamboa et al. (2021)** ainsi que **Manuel et al. (2021)** ont trouvé dans leurs travaux ou ils ont prouvé que la séquence de Pérou appartient à la lignée B.1.1.1 qui a subi une mutation dans le gène S.

Gong et al. (2020), montrent qu'il y a une délétion de 382 nucléotides dans le cadre de lecture ouvert 8 (ORF 8) qui a été trouvée dans une souche taïwanaise isolée d'un patient le 4 février 2020 qui avait des antécédents de voyage à Wuhan, ce que nous avons constaté dans l'arbre de gènes ORF8, où les séquences de Taiwan 1, Taiwan 2, Chine 3, Chine 5 et Chine 1 sont dans la même branche.

D'après les arbres des gènes N, ORF3a, ORF8 et ORF10 représentés dans les figures 53, 48, 64 et 61 respectivement, nous avons constaté que la Chine et la Colombie sont localisées dans la même branche pour chaque gène, chose qui a été prouvé par **Chu et al. (2021)** qui a montré que la quatrième vague épidémique de Cov 19 à Hong Kong a été introduite à partir de Colombie, par un voyageur de Colombie à Hong Kong.

L'arbre phylogénétique de gène S montre qu'il y a une liaison entre la séquence de Japon et celle du Brésil, cette liaison a été expliquée par **Fujino et al (2021)** qui ont trouvé que la mutation de S a été signalée au Japon, chez des personnes de retour du Brésil.

Pour revenir aux figures des arbres phylogénétiques des gènes S, N, nous remarquons que la Serbie prend des positions différentes, donc les gènes de la souche Serbienne sont souvent changés, selon **Miljanovic1 et al. (2021)** 148 mutations différentes ont été détectées en Serbie, la majorité touchent quatre gènes : ORF1ab, N, S et ORF8.

Dans les arbres phylogénétiques des gènes S, gène N, gène ORF3a, gène ORF8, gène ORF7b et gène ORF7a, nous trouvons que la position de la Tunisie est plus proche de l'Égypte, donc ont la même nature de souche, ce résultat confirme les travaux de **Laamarti et al. (2020)** qui a trouvé que toutes les souches de SARS Cov 2 d'Afrique du Nord portant la mutation D614G localisée sur gène S.

La comparaison des séquences des Etats Unis (USA) dans tous les arbres phylogénétiques, nous a permis de trouver que les positions de ces séquences dans les arbres et même dans les réseaux phylogénétiques pour chaque gène sont différentes, cela est expliqué dans les travaux de **Kaushal et al. (2020)** qui a montré que la majorité des gènes de l'USA accumulaient des mutations.

Conclusion

La pandémie Covid 19, a commencé en décembre 2019 à Wuhan en Chine, et a depuis touché de nombreux pays dans le monde, devenant une menace mondiale majeure. Le virus SARS Cov 2, appartient à la famille Coronaviridae, sous-famille Orthocorona virinae et au genre Betacoronavirus ; composé d'un ARN de grande taille qui porte 29 981 bases codant pour 9 860 acides aminés, conduisant à la production des protéines, utilise un récepteur d'entrée cellulaire, l'enzyme de conversion de l'angiotensine 2 (ACE2). L'organisation génomique de ce coronavirus est variable, il est composé par des gènes structurales qui sont les gènes : S, M, E et N et les gènes non structurales dits accessoires qui sont les gènes : ORF1ab, ORF6, ORF3a, ORF8, ORF10, ORF7a, ORF7b, ORFa1. Les analyses réalisées dans ce travail ont été élaborées à l'aide de différents logiciels entre autres, MEGA11 pour la reconstruction de l'arbre phylogénétique, DnaSP6 pour la détermination des haplotypes, Network pour avoir les réseaux, DAMBE et DnBA pour faire des calculs statistiques.

L'étude de la GenBank, nous a permis de classer 12 gènes différents de SARS Cov 2 à travers le monde qui sont : ORF1ab, ORF3a, N, S, ORF8, ORF10, ORF1a, ORF7a, ORF7b, ORF6, E, M ; les différents logiciels utilisés (MEGA, DnaSP, NetWork, DnaBA, DAMBE) nous ont permis de :

-Classer les gènes en 3 catégories, la 1^{ère} catégorie inclue les gènes les plus variables qui sont : ORF1ab, ORF3a, N, S. La 2^{ème} catégorie concerne les gènes les moins variables : ORF8, ORF10, ORF1a, ORF7a, ORF7b, et la 3^{ème} catégorie c'est pour les gènes non variés : ORF6, E, M.

-Les gènes les plus et les moins variables (ORF1ab, ORF3a, N, S, ORF8, ORF10, ORF1a, ORF7a, ORF7b) sont moins conservés, face aux transitions et transversions car ils contiennent un taux de AT élevé par rapport aux taux de CG.

-La comparaison entre les séquences de différents pays, par l'étude des positions de ces pays dans les arbres et les réseaux phylogénétiques, nous a prouvé que la souche de la Chine est à l'origine de toutes les autres souches du monde.

-Les résultats obtenus par Dnasp 6, DnBA et DAMBE qui montrent les différentes mutations qui touchent les gènes de SARS Cov 2 traduisent et confirment les résultats de MEGA11 et Network qui donnent des présentations graphiques des résultats obtenus sous forme d'arbres ou de réseaux phylogénétiques.

La propagation rapide de cette pandémie dans le monde est devenue une menace pour l'humanité, pour cela il y a un besoin urgent pour que les chercheurs s'unissent afin d'avancer rapidement dans leurs recherches sur ce virus pour le stopper.

Références Bibliographiques

- **Aguilar-Gamboa F. R., Salcedo-Mejía L.A., Serquén-López L.M., Mechallontop M.E., Tullume-Vergara P.O., Bonifacio-Briceño J.J, Salas-Asencios R., Silva Díaz H., Cárdenas J.P. 2021.** Genomic sequences and analysis of five SARS-CoV-2 variants obtained from patients in Lambayeque, Peru. *Microbiol Resour Announc* 10:e01267-20.
- **Allaby M., 2009.** A Dictionary of Zoology. Encyclopedia.com. Next Edition: 4 ed, Latest Edition (5 ed.). DOI: 10.1093/acref/9780199233410.001.0001.
- **Anderson J. D., Karel W. J., 2009.** A Genetic Assessment of Current Management Strategies for Spotted Seatrout in Texas, *Marine and Coastal Fisheries*, 1:1, 121-132, DOI: 10.1577/C09-001.1.
- **Aoues K., 2020.** Biodiagnostic du Covid 19. Communication, CHU Blida, EPH Tipaza.
- **Awadasseid A., Wu Y., Tanaka Y., Zhang W., 2021.** SARS-CoV-2 variants evolved during the early stage of the pandemic and effects of mutations on adaptation in Wuhan populations. *National Library of Medicine*, 1; 17(1):97-106. doi: 10.7150/ijbs.47827.
- **Baaziz M., 2017.** Phylogénie moléculaire basée sur le génome nucléaire, le plastome et le mitochondriome. <http://www.takween.com/techniques/phylogenie-moleculaire.html>
- **Basseur M., 2015.** **Introduction à la bioinformatique**. Département Informatique, Université d'Angers. P199.
- **Benalioua B., 2011.** Etude bioinformatique et dosage chimique des tanins produit par les plantes medicinales exposees au stress de la pollution. Mimosoir de magister spécialité : chimie. université Abdelhamid Ibn Badis Mostaganem.
- **Benhamdi A., 2020.** Théorie de l'évolution et phylogénie. Université des frères Mentouri Constantine 1.

- **Benson D., Mizrachi I., Lipman D.J., Ostell J., Wheeler D. L., 2008.** Nucleic Acids Research. Volume 36, Issue suppl_1, P: D25–D30
- **Chanthadavong A., 2020.** Covid-19 : IIA comme outil de diagnostic en hôpital. Fujitsu et l'hôpital Shinagawa de Tokyo.
- **Chu K.W. D., Hui P.Y. k., Gu H., Ko L. K.W, Krishnan P., Ng Y.M D., Liu Y.Z. G., Wan K. C. C., Chun Cheung M., Chun Ng k., Nicholls J., Tsang N.C.D., Peiris M., Chan C.W. M., Leo L.M., 2021.** Introduction of ORF3a-Q57H SARS-CoV-2 Variant Causing Fourth Epidemic Wave of COVID-19, Hong Kong, China. *Emerging Infectious Diseases*, 27(5), 1492-1495
- **Corre E., 2013.** Introduction aux méthodes de phylogénie. Formation Biogenouest, université de Rennes.
- **Deriham T., 2017.** Utilisation des arbres phylogénétiques dans l'alignement de séquence. Mémoire, université Mohamed Boudiaf - m'sila. P62. N° :84
- **Evans M., Hastings N., Peacock B., Felsenstein J., Johnson N.L., Kotz S., Kemp A.W., Wiley & Jukes S., Kimura M., Kuhner M.K., Felsenstein J., Lake J.A., Li W-H., Nee S., Holmes E.C., Rambaut A., Harvey P.H., Tamura K., Nei M., Xia X W-H., Yang Z., 2001.** DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. P: 371–373.
- **Fenstermacher D., 2005.** Introduction to bioinformatics. Journal de l'American Society for Information Science and Technology. Volume 56, Numéro 5. P: 440-446.
- **Forster P., Forster M., 2020.** Network Release notes. (https://www.fluxus-engineering.com/sharenet_rn.htm)
- **Fujino T., Nomoto H., Kutsuna S., Ujiie M., Suzuki T., Sato R., Fujimoto T., Kuroda M., Wakita T., Ohmagari N., 2021.** Novel SARS-CoV-2 Variant in Travelers from Brazil to Japan. *Emerg Infect Dis.* ; 27(4): 1243–1245. doi: 10.3201/eid2704.210138
- **Gala J. L., Nyabi O., François Durant J., Chibani N., Bentahir M., 2020.** Méthodes diagnostiques du COVID-19. *Biologie Et Technologies Moléculaires Appliquées*. P: 228-235.
- **Gattolliat J-L., 2002.** Etude systématique, cladistique et biogéographique des Baetidae (Ephemeroptera) de Madagascar. Thèse de doctorat en Zoologie et Ecologie Animale. Dir. Thèse : J-M Elouard. Univ. Lausanne. 111-145. 279p

- **Gong Y., Tsao K., Hsiao M., Huang C., Huang P. C., Huang P., Lee K., Liu Y., Yang S., Lin R., Chen K., Liu Y., Huang, Hsing-I S., Liu H., Yang J., Chiu C., Yang C., Chen G., Shih S., 2020.** SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerg Microbes Infect*; 9(1):1457-1466. doi: 10.1080/22221751.2020.1782271.
- **Graw J., 2006.** *Genetik*, Auflage, ISBN 978-3-540-29048-3. Springer Verlag Berlin Heidelberg.
- **Hennig W., 1975.** Cladistic Analysis or Cladistic Classification?: A Reply to Ernst Mayr. *Systematic Zoology*, 24(2): P244-256.
- **Hong Z, Yan F., Tao X., Wei-jian N., Al-zong S., Xiao-ming M., 2020.** Potential therapeutic targets and promising drugs for combating SARS-CoV-2. *Frère J Pharmacol*; 177(14):3147-3161. doi: 10.1111/bph.15092.
- **Hopkins S., 2021.** Categories: Coronavirus (Covid-19).
- **Hoppenot I., 2020.** Imagerie pulmonaire de la Covid-19 : Le scanner, un examen clé à prescrire à bon escient. VIDAL. L'intelligence medical au service du soin.
- **Huelsenbeck J. P., 1995.** The Robustness of Two Phylogenetic Methods: Four-Taxon Simulations Reveal a Slight Superiority of Maximum Likelihood over Neighbor Joining. *Mol. Biol. Evol.* 12(5): P: 843-849.
- **Huihui W., Xuemei L., Tao L., Shubing Z., Lianzi W., Xian W., Jiaqing L., 2020.** The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis.* P: 1–7.
- **Hunault G., Jaspard E., 2010.** LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics.* 1; 11: 221. doi: 10.1186/1471-2164-11-221.
- **Huson D. H., Bryant D., 2012.** Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*, 23(2):254-67. doi: 10.1093/molbev/msj030.
- **Jaspar E. Hunault G., 2016.** sHSPdb: a database for the analysis of small Heat Shock Proteins 135. *BMC Plant Biol.* 13;16(1):135. doi: 10.1186/s12870-016-0820-6.
- **Jawerth N., 2020.** Comment la RT-PCR en temps réel permet-elle de détecter le virus de la COVID-19 ?. IAEA: L'Agence internationale de l'énergie atomique. *Maladies infectieuses.* Volume: 61-2.

- **Kaushal N., Gupta Y., Goyal M., Khaiboullina S. F., Baranwal M., Verma S. C., 2020.** Mutational Frequencies of SARS-CoV-2 Genome during the Beginning Months of the Outbreak in USA. *P16*. 9(7), 565.
- **Korsia-Meffre S., 2021.** Covid-19 : mutations, variants, lignées, n501y, e484k... de quoi parle-t-on ?. *VIDAL. L'intelligence médicale au service du soin*.
- **Kourta D., 2021.** COVID-19 Deux cas ont été enregistrés à Alger : Le variant britannique détecté en Algérie. *El Waten*.
- **Koyama T., Platt D., Parida L., 2020.** Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ* 1; 98(7):495-504. doi: 10.2471/BLT.20.253591.
- **Laamarti M., Kartti S., Alouane T., Laamarti R., Allam L., Ouadghiri M., Chemaelfihri M.W., Smyej I., Rahoui J., Benrahma H., Diawara I., Essabbar A., Boumajdi N., Bendani H., Bouricha E., Aanniz T., Elattar J., EL Hafidi N., EL Jaoudi F., Sbabou L., Nejjari C., Amzazi S., Mentag R., Belyamani L, Ibrahimi A., 2020.** Genetic analysis of SARS-CoV-2 strains collected from North Africa: viral origins and mutational spectrum. *Genomic analysis of North African isolates*. P19.
- **Laib Ch., 2021.** Coronavirus : Ce qu'il faut savoir sur le variant nigérian. *Algerie 360°*.
- **Loayza-Alarico, Manuel J., De La Cruz-Vargas, Jhony A., 2021.** "Effect of SARS-CoV-2 variants on the transmission of COVID-19 in Peru," *Revista de la Facultad de Medicina Humana: Vol. 21: Iss. 1, Article 2*.
- **Mahalaxmi L., Kaavya J., Mohana D. S., Soo B. L., Ahmed A. D., Ssang-goo C., Balachandar V., 2020.** COVID-19: an update on diagnostic and therapeutic approaches. *BMB* ; P: 53(4):191–206. doi: 10.5483/BMBRep.2020.53.4.080.
- **Melika L. B., Michael R., Hamblinc D.E., Nima R., 2020.** COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clin Chim Acta*. 2020 Sep; 508: 254–266. doi: 10.1016/j.cca.2020.05.044
- **Mercatelli D., Giorgi F. M., 2020.** Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol*. 11:1800. doi: 10.3389/fmicb.2020.01800.
- **Miljanovic D., Milicevic O., Loncar A., Abazovic D., Despot D., Banko A., 2020.** The First Molecular Characterization of Serbian SARS-CoV-2 Isolates from a Unique Early Second Wave in Europe. *Front. Microbiol*. 12:691154. doi: 10.3389/fmicb.2021.691154

- **Moret B. M. E., Warnow T., 2002.** Reconstructing Optimal Phylogenetic Trees: A Challenge in Experimental Algorithmics. Lecture Notes in Computer Science, vol 2547. Springer, Berlin, Heidelberg. //doi.org/10.1007/3-540-36383-1_8
- **Moyou E., 2021.** Nombre de personnes guéries du coronavirus (COVID-19) dans le monde au 6 avril 2021, selon le pays. STATISTICA.
- **Patrick C. Y., Woo Susanna K. P., Lau. L. S. F.C., Lau C. Y., Tsang A. K. L. C., Lau J. H. N., Bai R., Teng L. L. J., Tsang C. C. C., Wang M., Zheng B., Chan K., Yuen K., 2020.** Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. Journal of Virology, vol. 86, no 7, p. 3995-4008 (PMID 22278237, DOI 10.1128/JVI.06540-11, Bibcode 3302495.
- **Philippe L., Didier C., Philippe H., 2002.** Molecular phylogeny and evolution. Bio-informatique (5). Med Sci (Paris) Volume 18, Number 11, P: 1146 – 1154
- **Plaçais L., Richier Q., 2020.** COVID-19: Clinical, biological and radiological characteristics in adults, infants and pregnant women. An up-to-date review at the heart of the pandemic. Rev Med Interne. 41(5):308-318. doi: 10.1016/j.revmed.2020.04.004.
- **Purva A., Gulam M., Sikhwinder S. S. M., Imtaiyaz H., 2020.** Molecular Basis of Pathogenesis of Coronaviruses: A Comparative Genomics Approach to Planetary Health to Prevent Zoonotic Outbreaks in the 21st Century. OMICS. 24(11):634-644. doi: 10.1089/omi.2020.0131.
- **Sukhadeo Baliram Barbuddhe, Deepak Bhiwa Rawool, a Pankaj Prakash Gaonkar, Jess Vergis, Kuldeep Dhama, Satyaveer Singh Malikd, 2020.** Global scenario, public health concerns and mitigation strategies to counter current ongoing SARS-CoV-2 / COVID-19 pandemic. Hum Vaccin Immunother. P: 1–11.
- **Richard A., Robertcd D., Valentinab D., Hodcroft E. B., Jan A., 2020.** Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. Swiss Med Wkly. 150 : w20224.
- **Rischkowsky B., Pilling D., 2008.** L'Etat des Ressources Zoogénétiques pour l'Alimentation et l'Agriculture dans le Monde. The Food and Agriculture Organization of the United Nations (FAO). P: 393-410.

- **Roy C., Mandal S. M., Mondal S. K., Mukherjee S., Mapder T., Ghosh W., Chakraborty R., 2020.** Trends of mutation accumulation across global SARS-CoV-2 genomes: Implications for the evolution of the novel Coronavirus. 2020050463 (doi: 10.20944/preprints202005.0463.v2).
- **Rozas J., Ferrer-Mata A., Carlos J., -DelBarrio A., Guirao-Rico S., Librado P., Sebastian E., Onsins R., Gracia A., 2021.** DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol*; 34 (12): 3299-3302. doi: 10.1093/molbev/msx248.
- **Rozhgar A. K., Safdarb M., Ozaslan M., 2020.** Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* 19:100682. doi: 10.1016/j.genrep.2020.100682.
- **Schmidt H-A., 2003-** Phylogenetic Trees from Large Datasets Inaugural. Thèse de Doctorat en Mathématique. Dir. Université de Heinrich–Heine–Düsseldorf vorgelegt von Heiko. 123p.
- **Sekulic M., Harper H., Nezami B. G., Shen L., Sekulic S. P., Koeth A. T., Harding V. C., Gilmore H., Sadri N., 2020.** Molecular Detection of SARS-CoV-2 Infection in FFPE Samples and Histopathologic Findings in Fatal SARS-CoV-2 Cases. *Am J Clin Pathol.*: aqaa091. doi: 10.1093/ajcp/aqaa091.
- **Sukhadeo B. B., Deepak B. R., Pankaj P. G., Jess V., 2020.** Global scenario, public health concerns and mitigation strategies to counter current ongoing SARS-CoV-2 / COVID-19 pandemic. *Hum Vaccin Immunother.* P: 1–11. doi: 10.1080/21645515.2020.1810496
- **Taguan C., 2015.** Les arbres phylogénétiques. Qu'est-ce que c'est ? Comment sont-ils construits ?. Programmation et algorithmique et Biologie.
- **Tillier E. R., Collins R. A., 2000.** The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *J Mol Evol.* 50(3):249-57. doi: 10.1007/s002399910029.
- **Umair M., Salman M., Rehman Z., Badar N., Ali Q., Abdul Ahad., Ikram A., 2021.** Proliferation of SARS-CoV-2 B.1.1.7 Variant in Pakistan-A Short Surveillance Account. *Front. Public Health* 9:683378. doi: 10.3389/fpubh.2021.683378
- **Wang J., Bao A., Gu J., He X., Wu Z., Qiao B., Chen Z., Xiong L., Zhang Y., Zheng H., Li F., Zhao Z., Mei S , Tong Y., 2021.** A retrospective analysis of 12,400 SARS-CoV-2 RNA tests in patients with COVID-19 in Wuhan. *Medicine (Baltimore).* 21; 100(20):e25916. doi: 10.1097/MD.00000000000025916

- **Xia X., Xie Z., 2001.** DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. Computer Notes 373. P3
- **Xia X., 2000.** Data Analysis in Molecular Biology and Evolution.
- **Xia X., 2013.** Department of Bio. DAMBE5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. Department of Biology and Center for Advanced Research in Environmental Genomics. University of Ottawa, Ottawa, Ontario, Canada. P9.
- **Yaeghobi R., 2017.** Utilisation des arbres phylogénétiques dans l'alignement de séquence. Mémoire présenté pour l'obtention du diplôme de Master Académique, université Mohamed Boudiaf - M'sila, département : Informatique.
- **Ye Z., Yuan S., 2020.** Zoonotic origins of human coronaviruses. Int J Biol Sci ; 16(10):1686-1697. doi:10.7150/ijbs.45472
- **Yoann M., 2012.** Les arbres phylogénétiques : construction et interprétation. Bioinf-fr.net Geekus biologicus.
- **Zi-wei Y., Shuofeng Y., Kit-san Y., Sin-ye F., Chi-ping C., Dong-yan J., 2020.** Zoonotic origins of human coronaviruses. Int J Biol Sci. ; 16(10): 1686–1697. doi: 10.7150/ijbs.45472

Biblio-Web

- **OMS:** <https://www.who.int/fr/>
- **NCBI:** <https://www.ncbi.nlm.nih.gov/genbank/>
- **FABOX :** <https://users-birc.au.dk/palle/php/fabox/>
- **FAS to RDF Conversion :** <http://fas2rdf.herokuapp.com/>
- **MEGA:** <https://www.megasoftware.net/>
- **DAMBE :** <http://dambe.bio.uottawa.ca/DAMBE/dambe.aspx>
- **DnaSP :** <http://www.ub.edu/dnasp/>
- **Network:** <https://www.fluxus-engineering.com/sharenet.htm>

Nous avons présenté dans cette partie pour chaque gène des 12 gènes de Sars Cov 2 étudiés (ORF1ab, S, N, ORF3a, ORF7a, ORF7b, ORF8, ORF10, ORFa1, M, ORF6, E) : **1.** l'arbre phylogénétique obtenu par MEGA, **2.** le réseau phylogénétique de Network, **3.** l'analyse statistique de DnBA, **4.** l'analyse graphique de DAMBE cela dans les mesures du possible.

1. Les gènes les plus variables:

1.1. Gene S



Figure 44 : Arbre phylogénétique du gène S de SARS Cov 2 construit à partir de 37 séquences d'ADN, en utilisant la méthode UPGMA par logiciel MEGA.

➤ Les résultats de DnBA pour le gène S sont représentés dans le chapitre III figure 38.

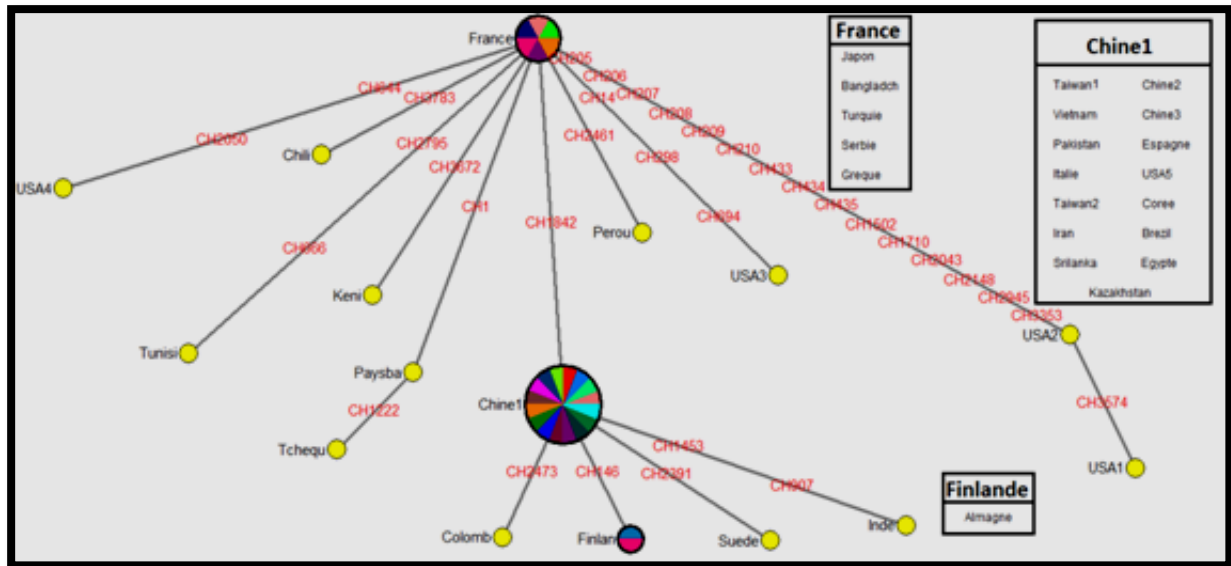
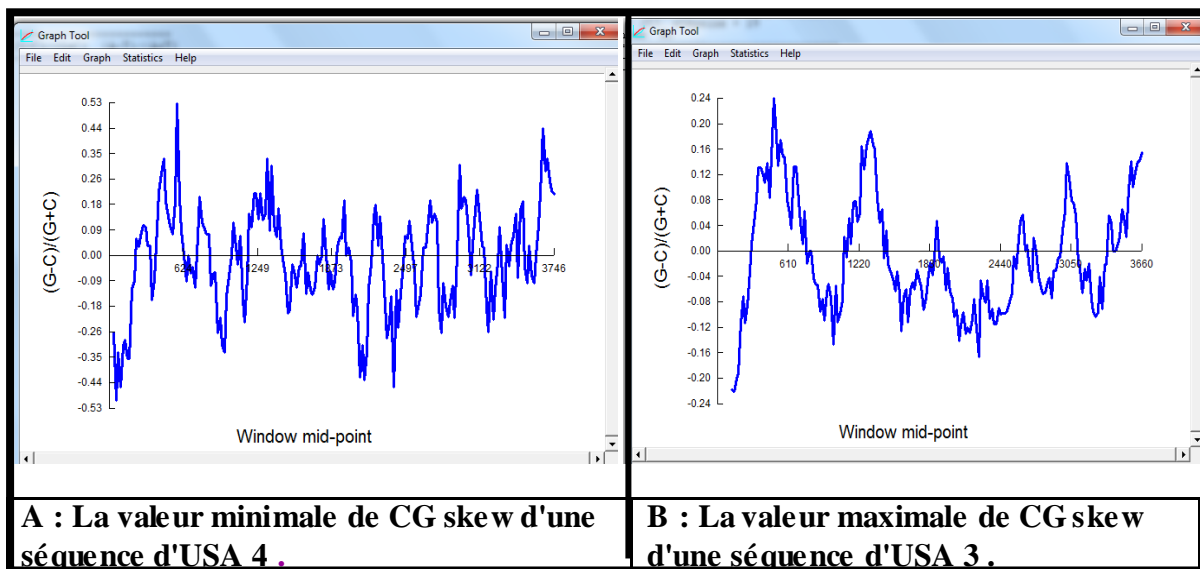


Figure 45 : Réseau phylogénétique de séquences de Gene S de SARS Cov 2, reconstruit par le logiciel NetWork à partir de 37 séquences, représentées par des cercles colorés, chaque couleur représente un pays.



A : La valeur minimale de CG skew d'une séquence d'USA 4 .

B : La valeur maximale de CG skew d'une séquence d'USA 3 .

Figure 46 : Variations du GC skew de gène S réalisées par le logiciel DAMBE.

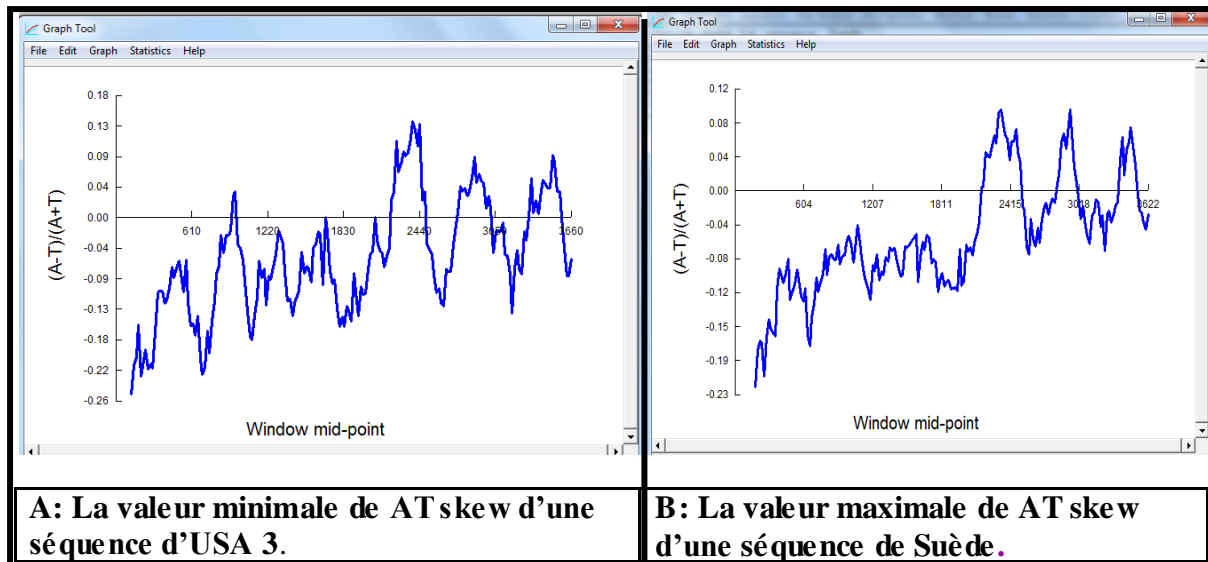


Figure 47 : Variations du AT skew de gène S réalisées par le logiciel DAMBE.

1.2. Gene ORF3a

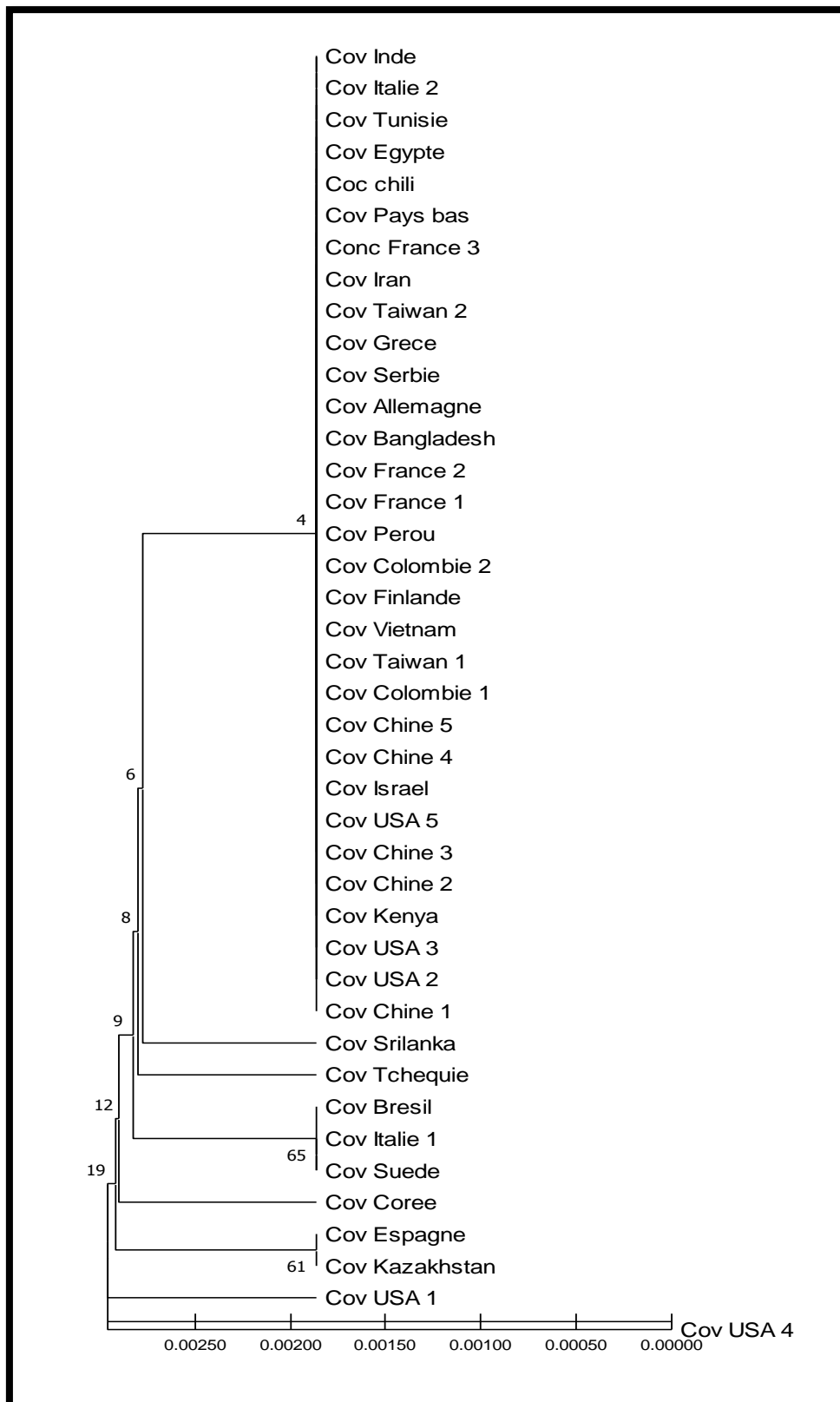


Figure 48 : Arbre phylogénétique de gène ORF3a de SARS Cov 2 construit à partir de 41 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

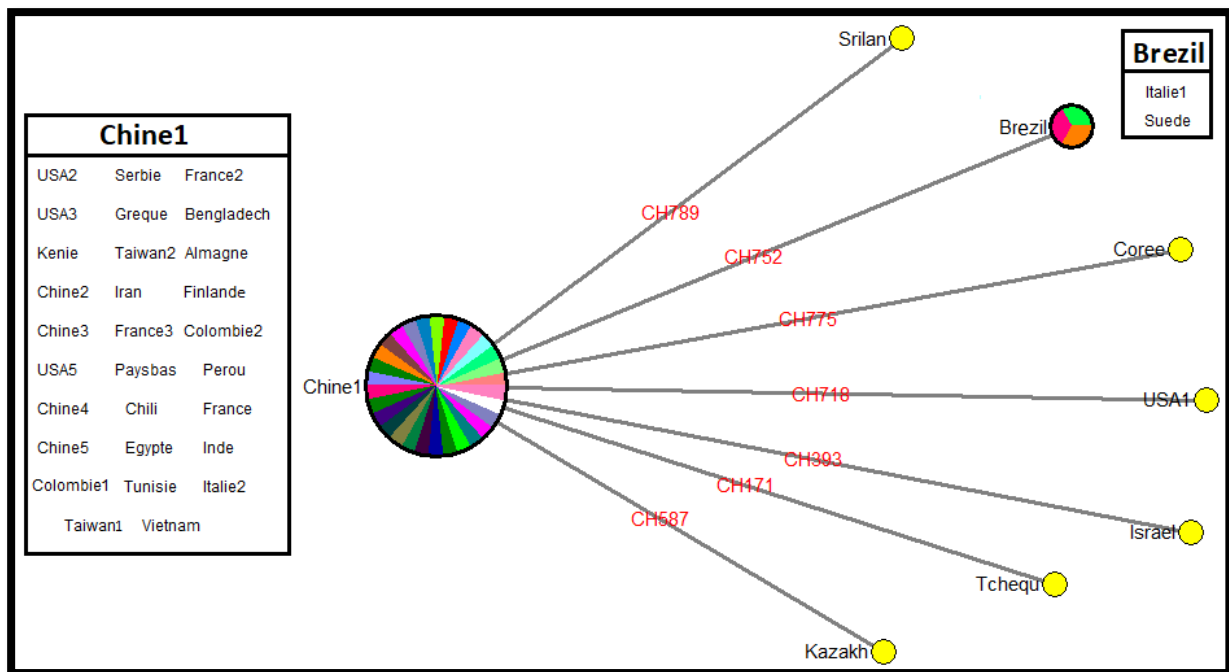


Figure 49 : Réseau phylogénétique de séquences de gène ORF3a de SARS Cov2, reconstruit par le logiciel NetWork à partir de 41 séquences, représentées par des cercles colorés, chaque couleur représente un pays.

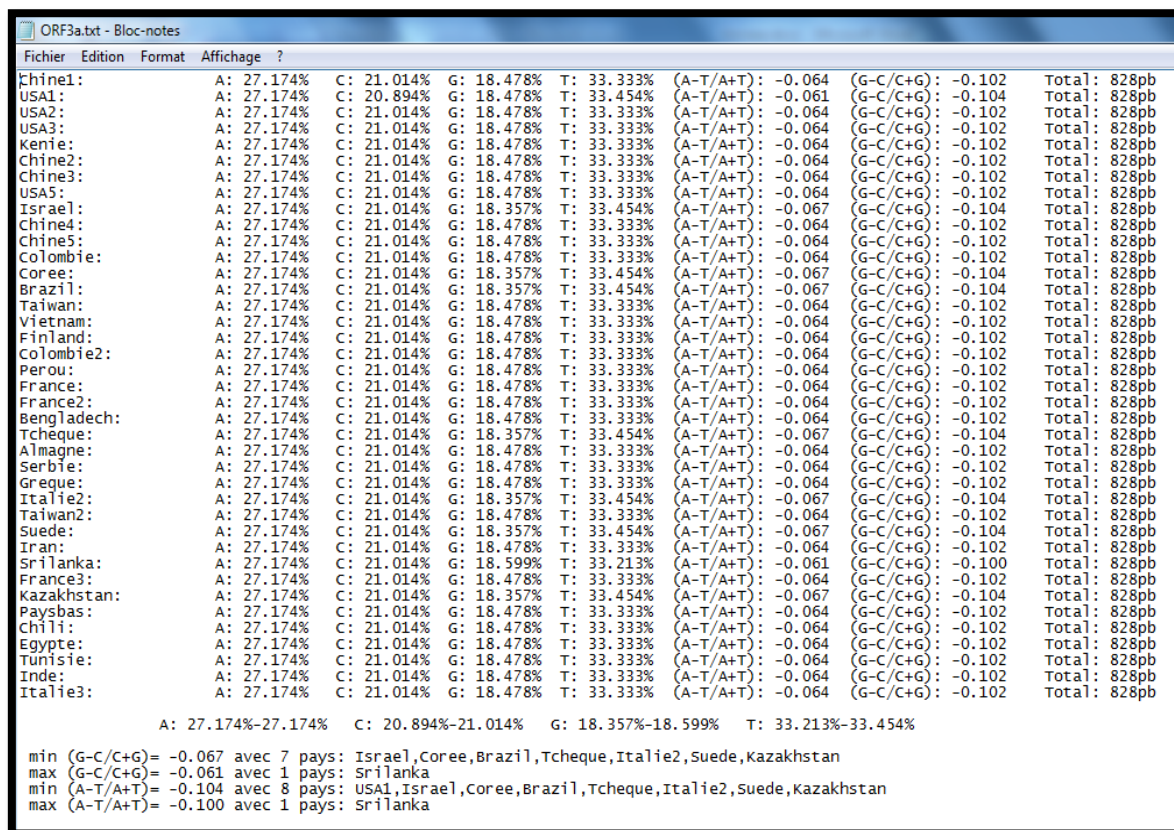


Figure 50: Analyse des bases azotiques des séquences de gène ORF3a par DnBA.

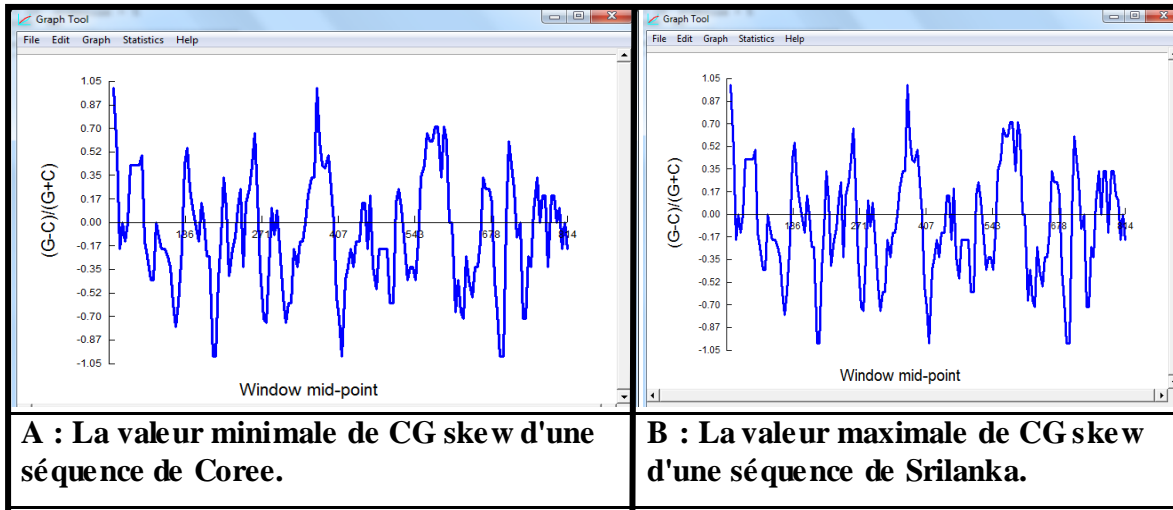


Figure 51: Variations du GC skew de gène ORF3a réalisées par le logiciel DAMBE.

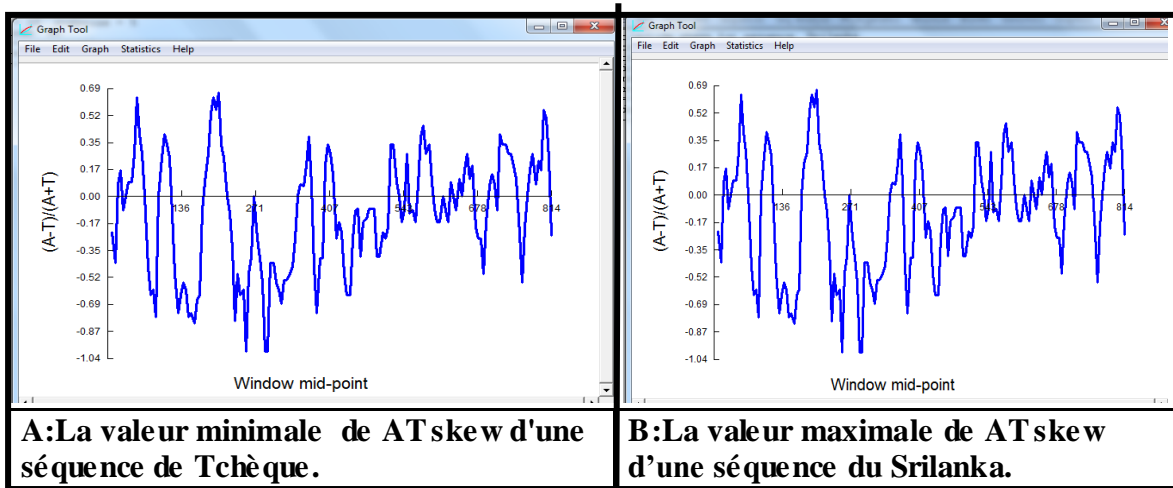


Figure 52: Variations de AT skew de gène ORF3a réalisées par logiciel DAMBE.

1.3. Gène N

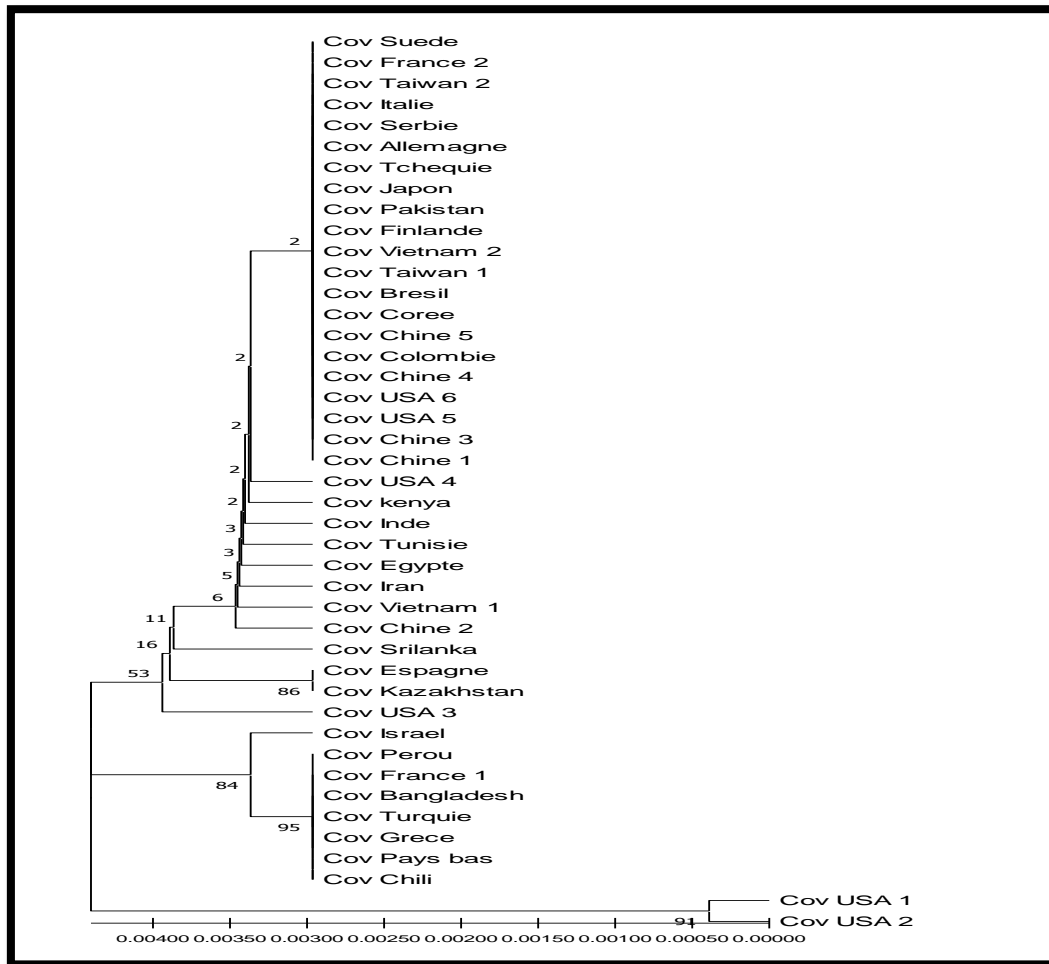


Figure 53 : Arbre phylogénétique de gène N de SARS Cov 2 construit à partir de 43 séquences, en utilisant la méthode UPGMA par logiciel MEGA.

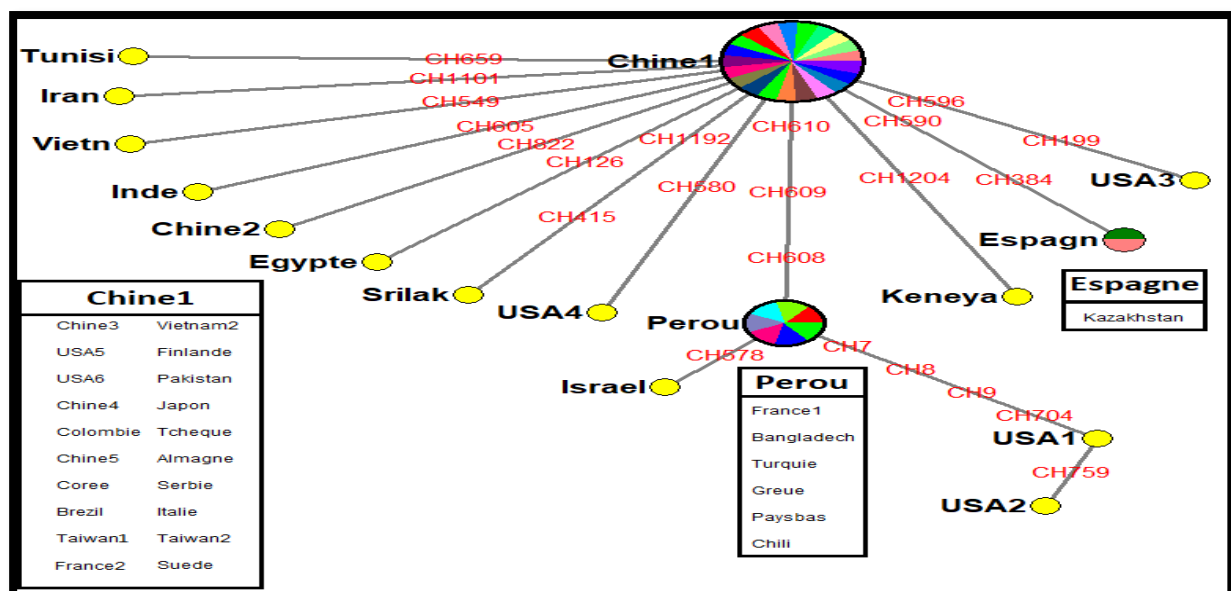


Figure 54 : Réseau phylogénétique de séquences de Gene N de SARS Cov2, reconstruit par le logiciel NetWork à partir de 43 séquences, représentées par des cercles colorés, chaque couleur représente un pays.

Fichier	Edition	Format	Affichage ?	A	C	G	T	(A-T)/(A+T)	(G-C)/(G+C)	Total				
Chine1:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
USA1:	A:	31.905%	C:	25.079%	G:	21.905%	T:	21.111%	(A-T)/(A+T):	-0.068	(G-C)/(G+C):	0.204	Total:	1260pb
USA2:	A:	31.984%	C:	25.079%	G:	21.825%	T:	21.111%	(A-T)/(A+T):	-0.069	(G-C)/(G+C):	0.205	Total:	1260pb
USA3:	A:	31.746%	C:	24.841%	G:	22.222%	T:	21.190%	(A-T)/(A+T):	-0.056	(G-C)/(G+C):	0.199	Total:	1260pb
Kenie:	A:	31.746%	C:	25.000%	G:	22.143%	T:	21.111%	(A-T)/(A+T):	-0.061	(G-C)/(G+C):	0.201	Total:	1260pb
USA4:	A:	31.746%	C:	25.000%	G:	22.302%	T:	20.952%	(A-T)/(A+T):	-0.057	(G-C)/(G+C):	0.205	Total:	1260pb
Chine2:	A:	31.746%	C:	24.921%	G:	22.222%	T:	21.111%	(A-T)/(A+T):	-0.057	(G-C)/(G+C):	0.201	Total:	1260pb
Chine3:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Espagne:	A:	31.746%	C:	24.841%	G:	22.222%	T:	21.190%	(A-T)/(A+T):	-0.056	(G-C)/(G+C):	0.199	Total:	1260pb
USA5:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Israël:	A:	31.905%	C:	25.079%	G:	21.905%	T:	21.111%	(A-T)/(A+T):	-0.068	(G-C)/(G+C):	0.204	Total:	1260pb
USA6:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Chine4:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Colombie:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Chine5:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Coree:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Vietn:	A:	31.746%	C:	25.079%	G:	22.222%	T:	20.952%	(A-T)/(A+T):	-0.060	(G-C)/(G+C):	0.205	Total:	1260pb
Bresil:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Taiwan 1:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Vietnam2:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Finlande:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Pakistan:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Perou:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
France1:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
Japon:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Bangladesh:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
Tcheque:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Almagne:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Turquie:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
Serbie:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Greue:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
Italie1:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Taiwan2:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Suede:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Iran:	A:	31.825%	C:	25.000%	G:	22.143%	T:	21.032%	(A-T)/(A+T):	-0.061	(G-C)/(G+C):	0.204	Total:	1260pb
Srilaka:	A:	31.746%	C:	25.079%	G:	22.143%	T:	21.032%	(A-T)/(A+T):	-0.062	(G-C)/(G+C):	0.203	Total:	1260pb
France2:	A:	31.746%	C:	25.000%	G:	22.222%	T:	21.032%	(A-T)/(A+T):	-0.059	(G-C)/(G+C):	0.203	Total:	1260pb
Kazakhstan:	A:	31.746%	C:	24.841%	G:	22.222%	T:	21.190%	(A-T)/(A+T):	-0.056	(G-C)/(G+C):	0.199	Total:	1260pb
Paysbas:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
Chili:	A:	31.905%	C:	25.079%	G:	21.984%	T:	21.032%	(A-T)/(A+T):	-0.066	(G-C)/(G+C):	0.205	Total:	1260pb
Egypte:	A:	31.746%	C:	24.921%	G:	22.222%	T:	21.111%	(A-T)/(A+T):	-0.057	(G-C)/(G+C):	0.201	Total:	1260pb
Tunisie:	A:	31.746%	C:	24.921%	G:	22.222%	T:	21.111%	(A-T)/(A+T):	-0.057	(G-C)/(G+C):	0.201	Total:	1260pb
Inde:	A:	31.825%	C:	25.000%	G:	22.143%	T:	21.032%	(A-T)/(A+T):	-0.061	(G-C)/(G+C):	0.204	Total:	1260pb

A: 31.746%-31.984% C: 24.841%-25.079% G: 21.825%-22.302% T: 20.952%-21.190%

min (G-C)/(G+C) = -0.069 avec 1 pays: USA2
max (G-C)/(G+C) = -0.056 avec 3 pays: USA3, Espagne, Kazakhstan
min (A-T)/(A+T) = 0.199 avec 3 pays: USA3, Espagne, Kazakhstan
max (A-T)/(A+T) = 0.205 avec 7 pays: Perou, France1, Bangladesh, Turquie, Greue, Paysbas, Chili

Figure 55 : Analyse des bases azotiques des séquences de gène N par DnBA.

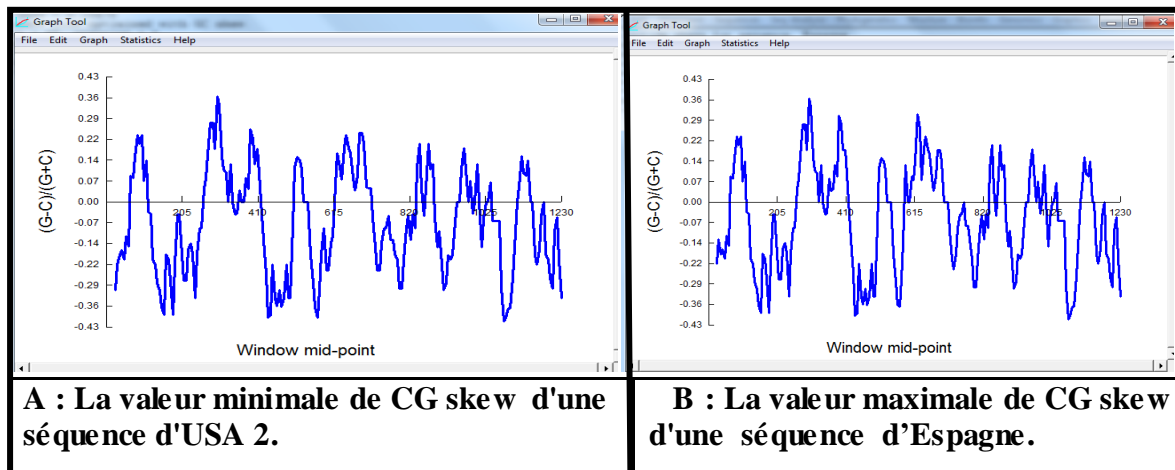


Figure 56 : Variations du GC skew de gène N réalisées par le logiciel DAMBE.

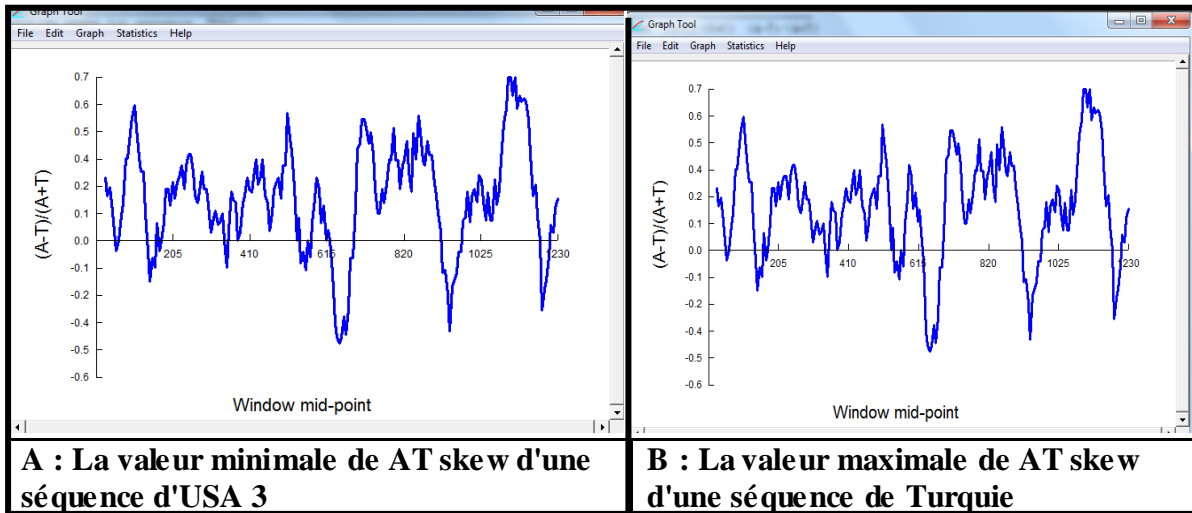


Figure 57: Variations du AT skew de gène N réalisées par le logiciel DAMBE.

1.4. Gène ORF1ab

Fichier	Edition	Format	Affichage	?
Bresil:	A: 30.075%	C: 17.745%	G: 19.915%	T: 32.265%
Bangladesh:	A: 30.070%	C: 17.735%	G: 19.915%	T: 32.280%
Egypte:	A: 30.080%	C: 17.745%	G: 19.915%	T: 32.260%
Coree:	A: 30.070%	C: 17.745%	G: 19.920%	T: 32.265%
Vietnam:	A: 30.075%	C: 17.740%	G: 19.920%	T: 32.265%
USA5:	A: 30.080%	C: 17.740%	G: 19.920%	T: 32.260%
USA4:	A: 30.075%	C: 17.730%	G: 19.900%	T: 32.295%
USA3:	A: 30.070%	C: 17.715%	G: 19.915%	T: 32.300%
USA2:	A: 30.089%	C: 17.703%	G: 19.919%	T: 32.290%
USA1:	A: 30.084%	C: 17.708%	G: 19.929%	T: 32.280%
Turquie:	A: 30.075%	C: 17.740%	G: 19.920%	T: 32.265%
Tunisie:	A: 30.077%	C: 17.726%	G: 19.901%	T: 32.297%
Taiwan2:	A: 30.070%	C: 17.740%	G: 19.925%	T: 32.265%
Taiwan1:	A: 30.075%	C: 17.745%	G: 19.920%	T: 32.260%
Suede:	A: 30.070%	C: 17.745%	G: 19.930%	T: 32.255%
Serbie:	A: 30.075%	C: 17.735%	G: 19.920%	T: 32.270%
Perou:	A: 30.070%	C: 17.735%	G: 19.925%	T: 32.270%
Paysbas:	A: 30.075%	C: 17.735%	G: 19.920%	T: 32.270%
Pakistan:	A: 30.080%	C: 17.730%	G: 19.910%	T: 32.280%
Colombie:	A: 30.075%	C: 17.745%	G: 19.920%	T: 32.260%
Kazakhstan:	A: 30.085%	C: 17.735%	G: 19.915%	T: 32.265%
Italie:	A: 30.070%	C: 17.745%	G: 19.920%	T: 32.265%
Israël:	A: 30.080%	C: 17.710%	G: 19.910%	T: 32.300%
Iran:	A: 30.080%	C: 17.740%	G: 19.910%	T: 32.270%
Inde:	A: 30.075%	C: 17.730%	G: 19.920%	T: 32.275%
Greque:	A: 30.075%	C: 17.735%	G: 19.920%	T: 32.270%
France2:	A: 30.070%	C: 17.730%	G: 19.920%	T: 32.275%
France1:	A: 30.075%	C: 17.740%	G: 19.920%	T: 32.265%
Espagne:	A: 30.080%	C: 17.730%	G: 19.920%	T: 32.270%
Finlande:	A: 30.075%	C: 17.745%	G: 19.920%	T: 32.260%
Chine5:	A: 30.075%	C: 17.745%	G: 19.920%	T: 32.260%
Chine4:	A: 30.075%	C: 17.730%	G: 19.920%	T: 32.275%
Chine3:	A: 30.075%	C: 17.745%	G: 19.920%	T: 32.260%
Chine2:	A: 30.075%	C: 17.740%	G: 19.920%	T: 32.265%
Chine1:	A: 30.070%	C: 17.745%	G: 19.930%	T: 32.255%
Chili:	A: 30.075%	C: 17.735%	G: 19.920%	T: 32.270%
Almagne:	A: 30.083%	C: 17.742%	G: 19.917%	T: 32.258%

A: 30.070%-30.089%	C: 17.703%-17.745%	G: 19.900%-19.930%	T: 32.255%-32.300%
--------------------	--------------------	--------------------	--------------------

min (G-C/C+G) = 0.058 avec 2 pays: Bresil,Egypte
max (G-C/C+G) = 0.059 avec 1 pays: USA1
min (A-T/A+T) = -0.036 avec 1 pays: USA3
max (A-T/A+T) = -0.035 avec 1 pays: Almagne

Figure 58: Analyse des bases azotiques des séquences de gène ORF1ab par DnBA.

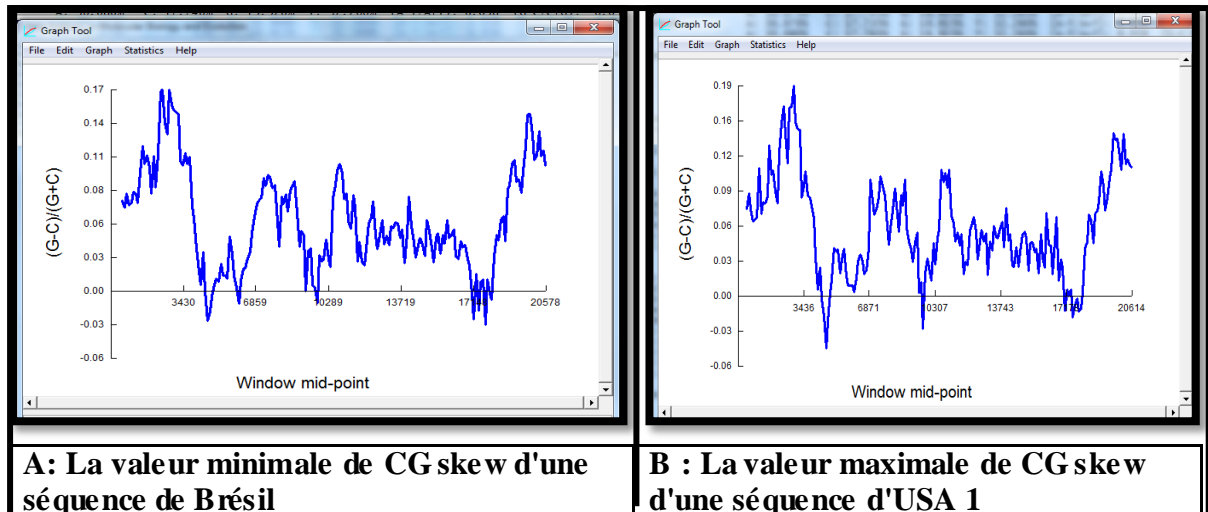


Figure 59 : Variations du GC skew de gène ORF1ab réalisées par le logiciel DAMBE.

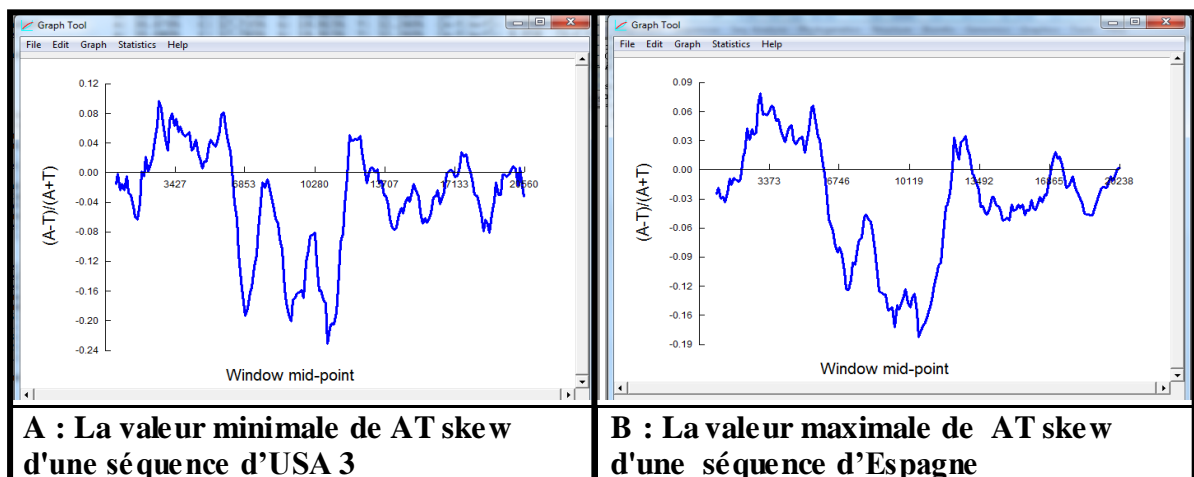


Figure 60 : Variations du AT skew de gène ORF1ab réalisées par le logiciel DAMBE.

- Les résultats de MEGA et NetWork pour le gène ORF1ab sont représentés respectivement dans les figures 33, 42 dans le chapitre III.

2. Les gènes moins variables

2.1. Gène ORF10

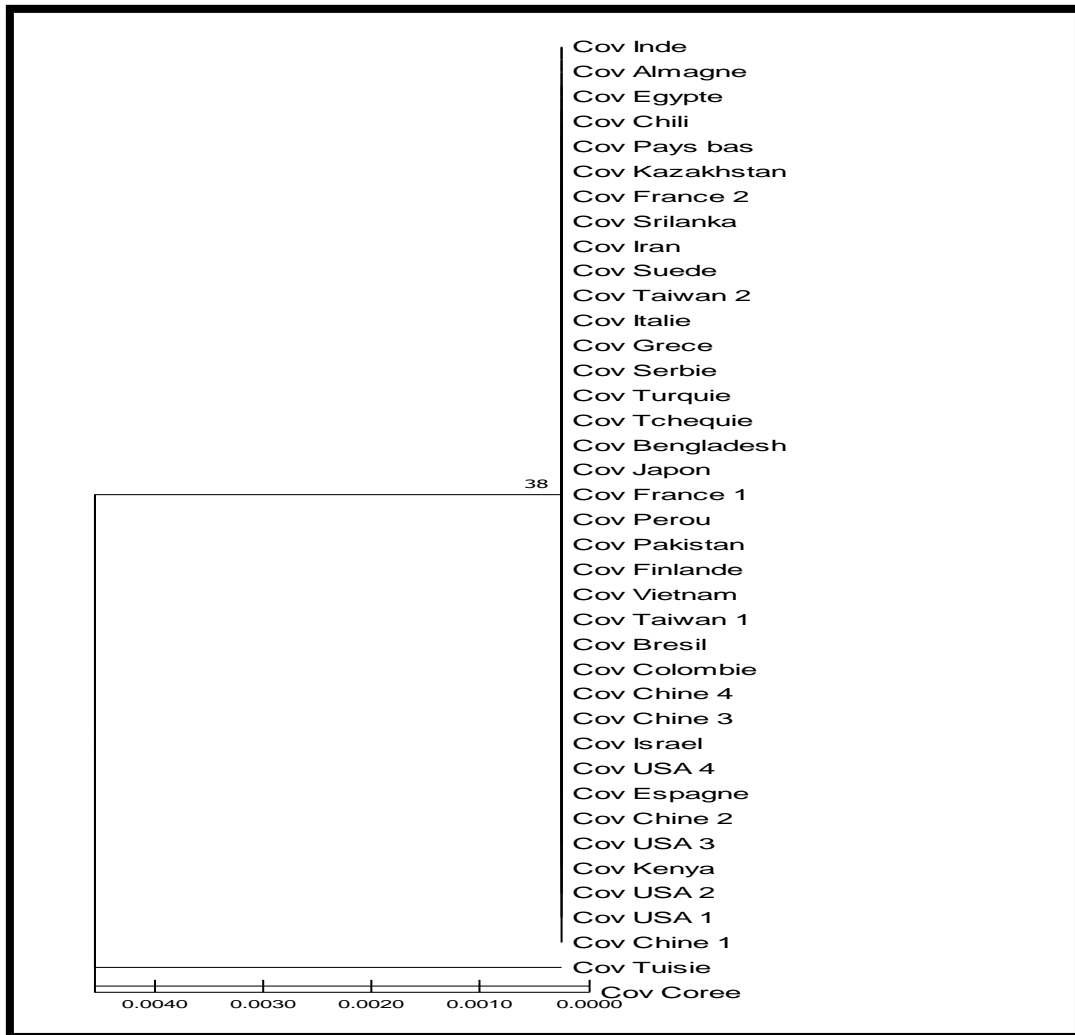


Figure 61 : Arbre phylogénétique de gène ORF10 de SARS Cov 2 construit à partir de 39 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

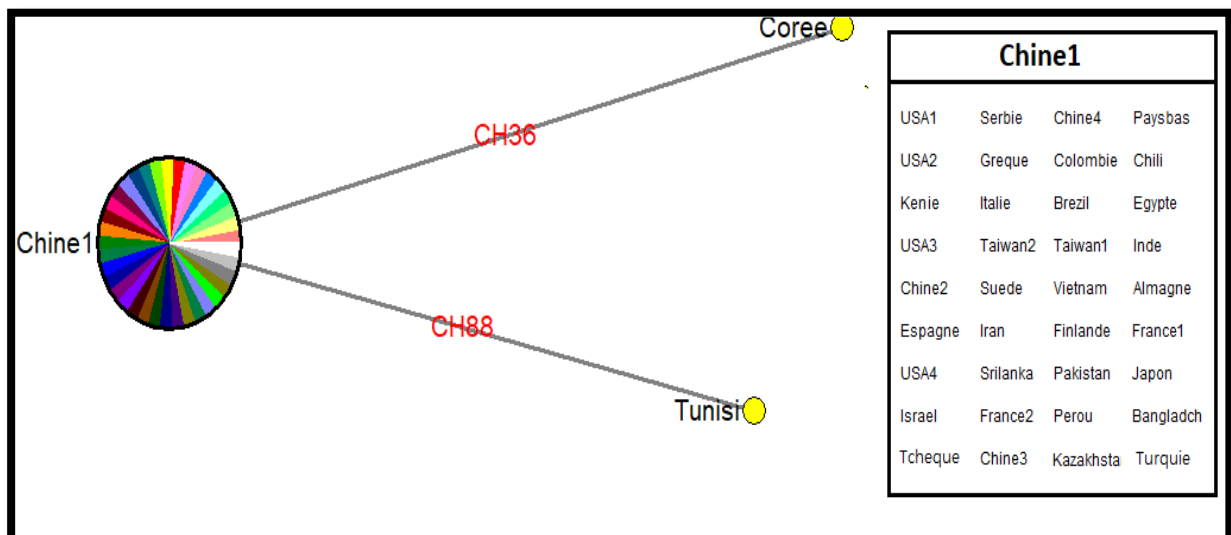


Figure 62 : Réseau phylogénétique de séquences de Gene ORF10 de SARS Cov 2, reconstruit par le logiciel NetWork à partir de 39 séquences, représentées par des cercles colorés, chaque couleur représente un pays.

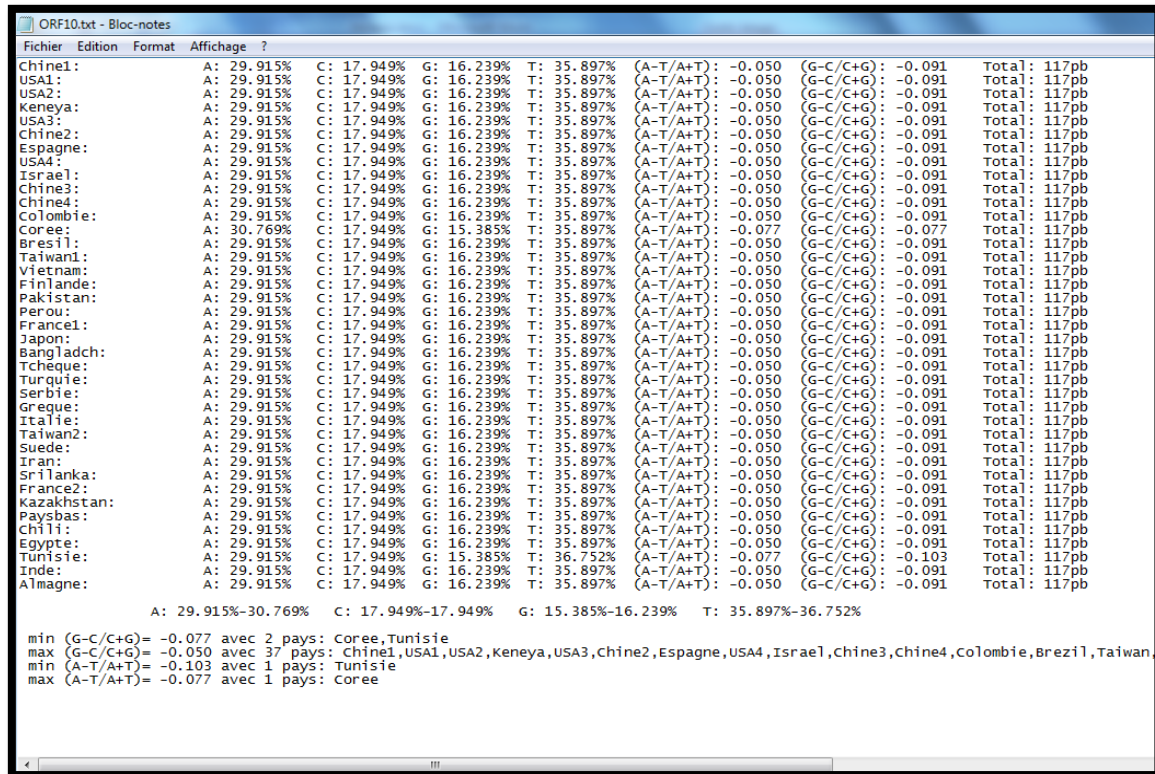


Figure 63 : Analyse des bases azotiques des séquences de gène ORF10 par DnBA.

2.2. Gene ORF8

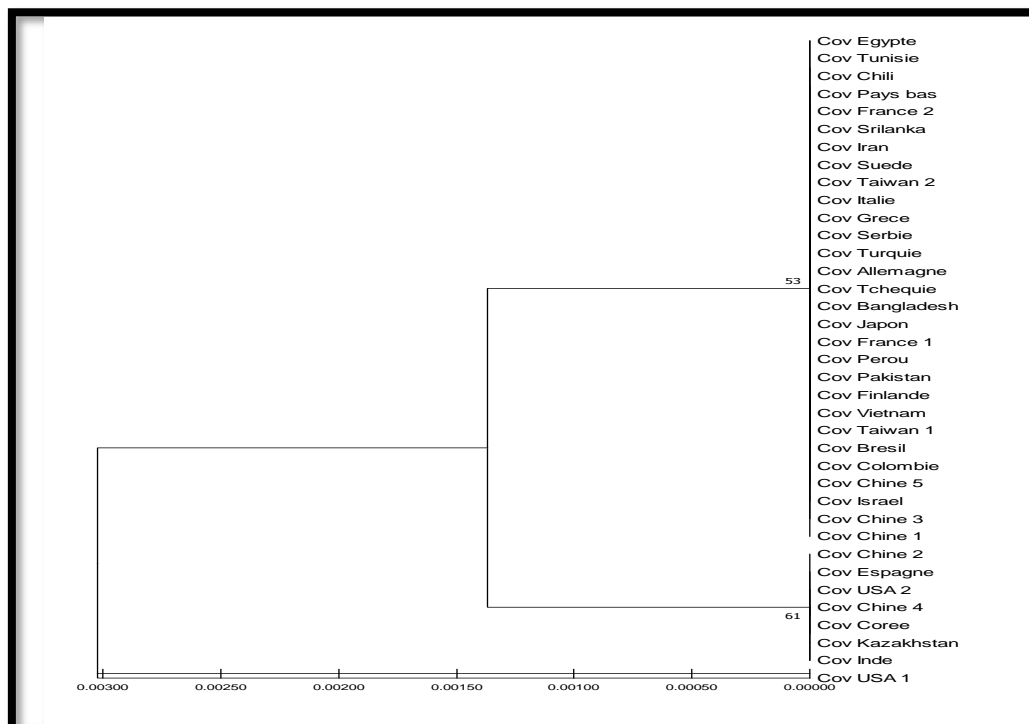


Figure 64 : Arbre phylogénétique de gène ORF8 de SARS Cov 2 construit à partir de 37 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

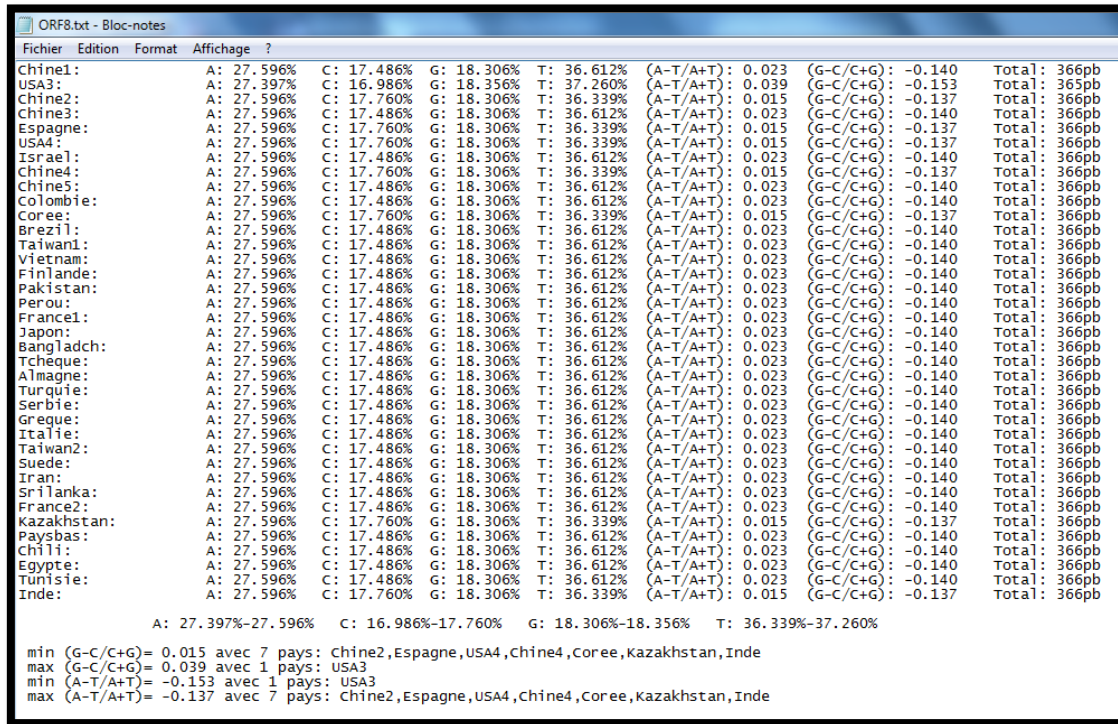


Figure 65 : Analyse des bases azotiques des séquences de gène ORF8 par DnBA.

2.3. Gène ORF7b

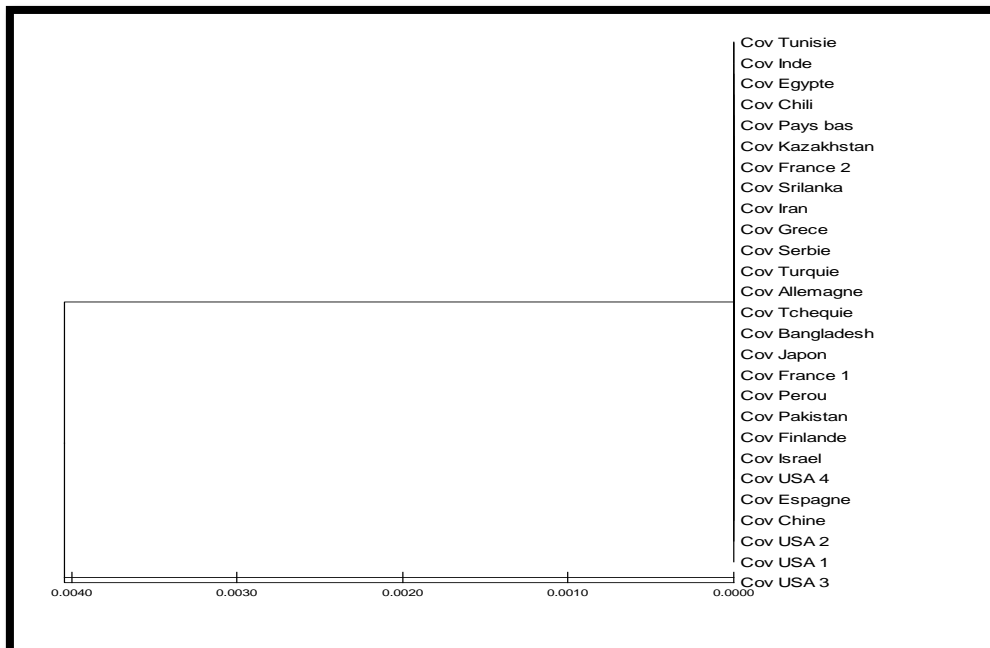


Figure 66 : Arbre phylogénétique de gène ORF7b de SARS Cov 2 construit à partir de 26 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

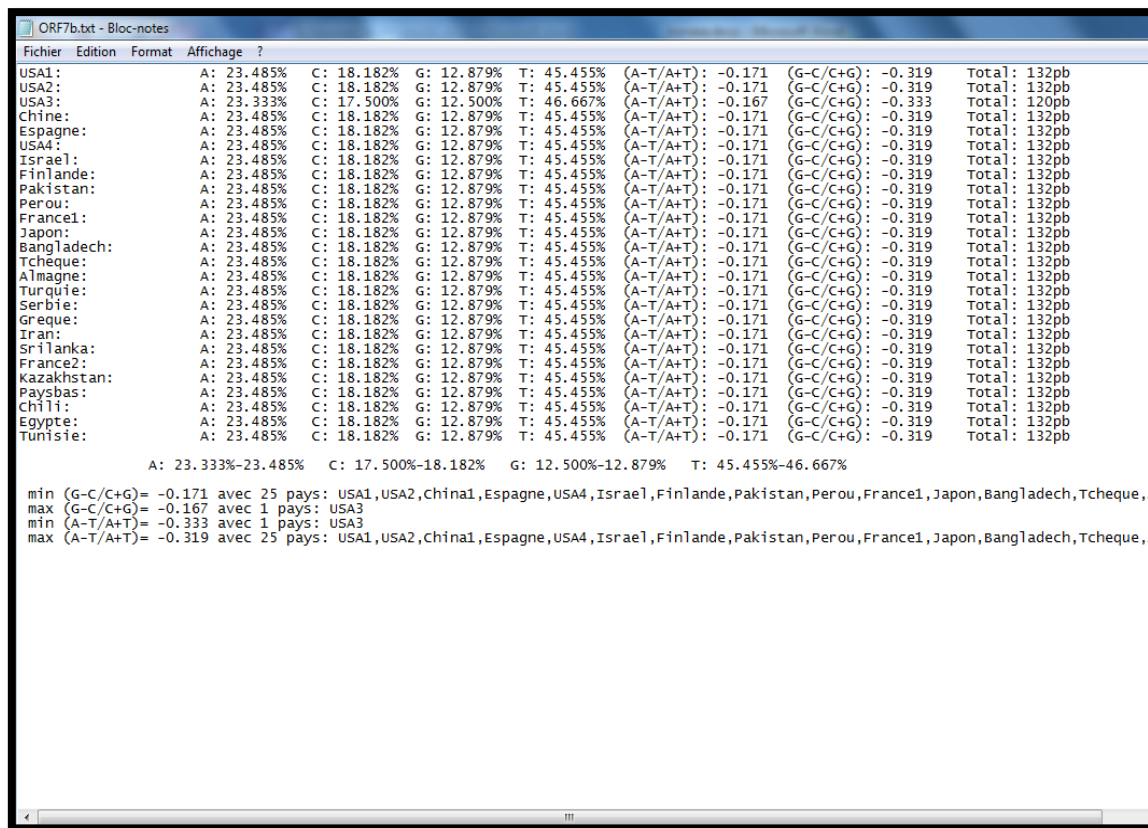


Figure 67 : Analyse des bases azotiques des séquences de gène ORF7b par DnBA.

2.4. Gène ORFa1

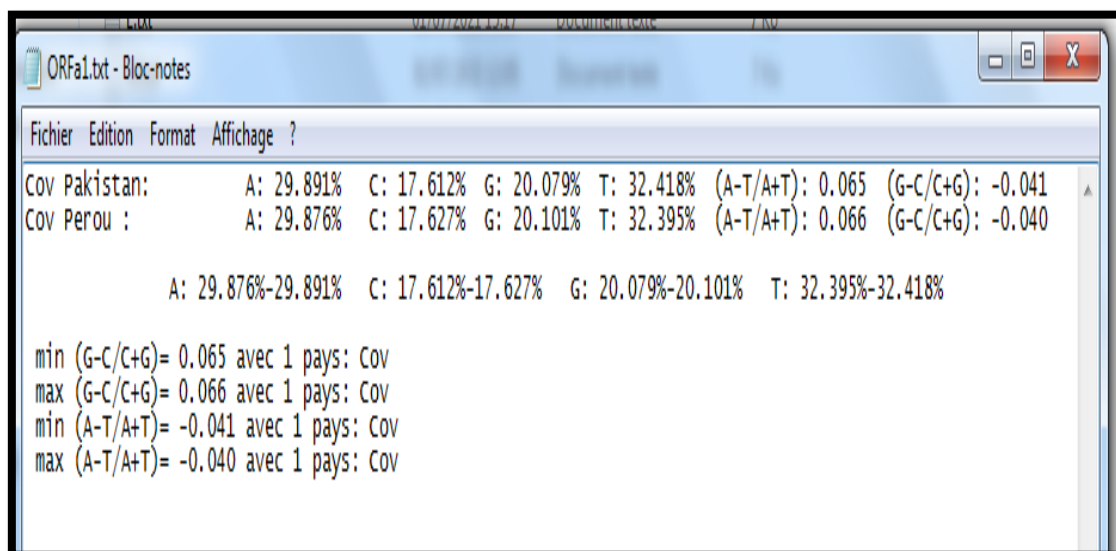


Figure 68 : Analyse des bases azotiques des 2 séquences de gène ORFa1 par DnBA.

2.5. Gène ORF7a

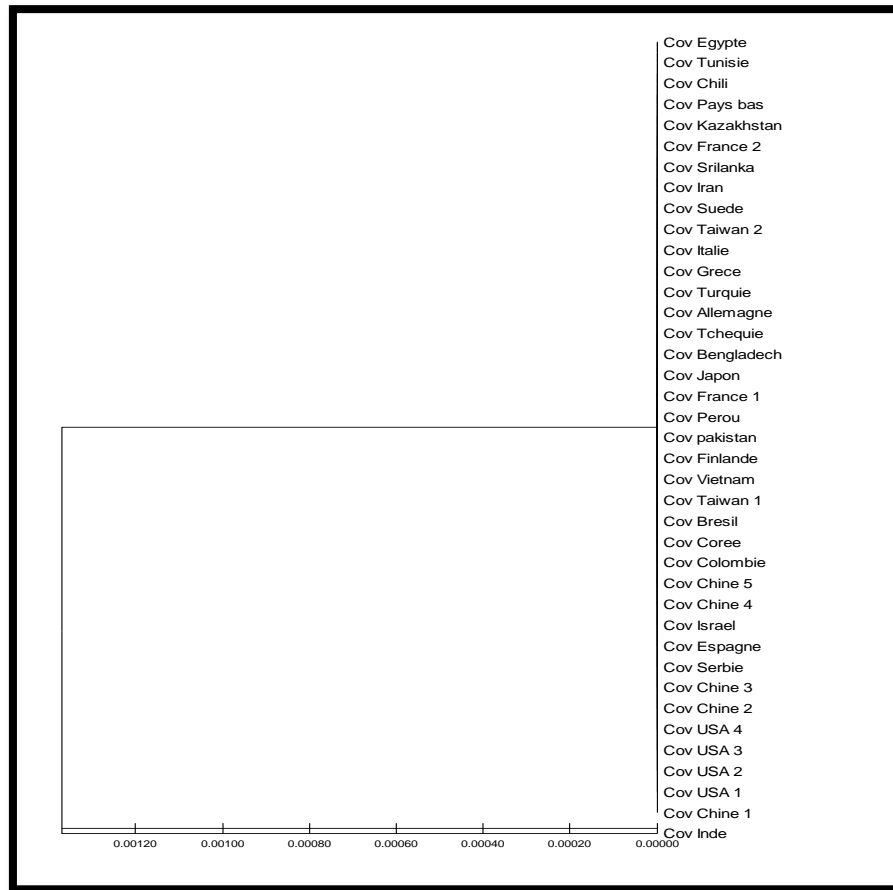


Figure 69 : Arbre phylogénétique de gène ORF7a de SARS Cov 2 construit à partir de 39 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

ORF7a.txt - Bloc-notes									
Fichier	Edition	Format	Affichage ?						
Chine1:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
USA1:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
USA2:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
USA3:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
USA4:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Chine2:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Chine3:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Serbie:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Espagne:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Israël:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Chine4:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Chine5:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Colombie:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Coree:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Brezil:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Taiwan:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Vietnam:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Finlande:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Pakistan:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Perou:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
France1:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Japon:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Bangladch:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Tcheque:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Allemagne:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Turquie:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Grece:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Italie:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Taiwan2:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Suede:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Iran:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Sri Lanka:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
France2:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Kazakhstan:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Paysbas:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Chili:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Egypte:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Tunisie:	A: 29.508%	C: 21.585%	G: 16.667%	T: 32.240%	(A-T/A+T): -0.129	(G-C/C+G): -0.044	Total:	366pb	
Inde:	A: 29.235%	C: 21.585%	G: 16.940%	T: 32.240%	(A-T/A+T): -0.121	(G-C/C+G): -0.049	Total:	366pb	
A: 29.235%-29.508% C: 21.585%-21.585% G: 16.667%-16.940% T: 32.240%-32.240%									
min (G-C/C+G) = -0.129 avec 38 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,serbie,Espagne,Israël,Chine4,Chine5,Colombie,Coree,									
max (G-C/C+G) = -0.121 avec 1 pays: Inde									
min (A-T/A+T) = -0.049 avec 1 pays: Inde									
max (A-T/A+T) = -0.044 avec 38 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,serbie,Espagne,Israël,Chine4,Chine5,Colombie,Coree,									

Figure 70 : Analyse des bases azotiques des séquences de gène ORF7a par DnBA.

3 .Les gènes non variables

3.1. Gene E

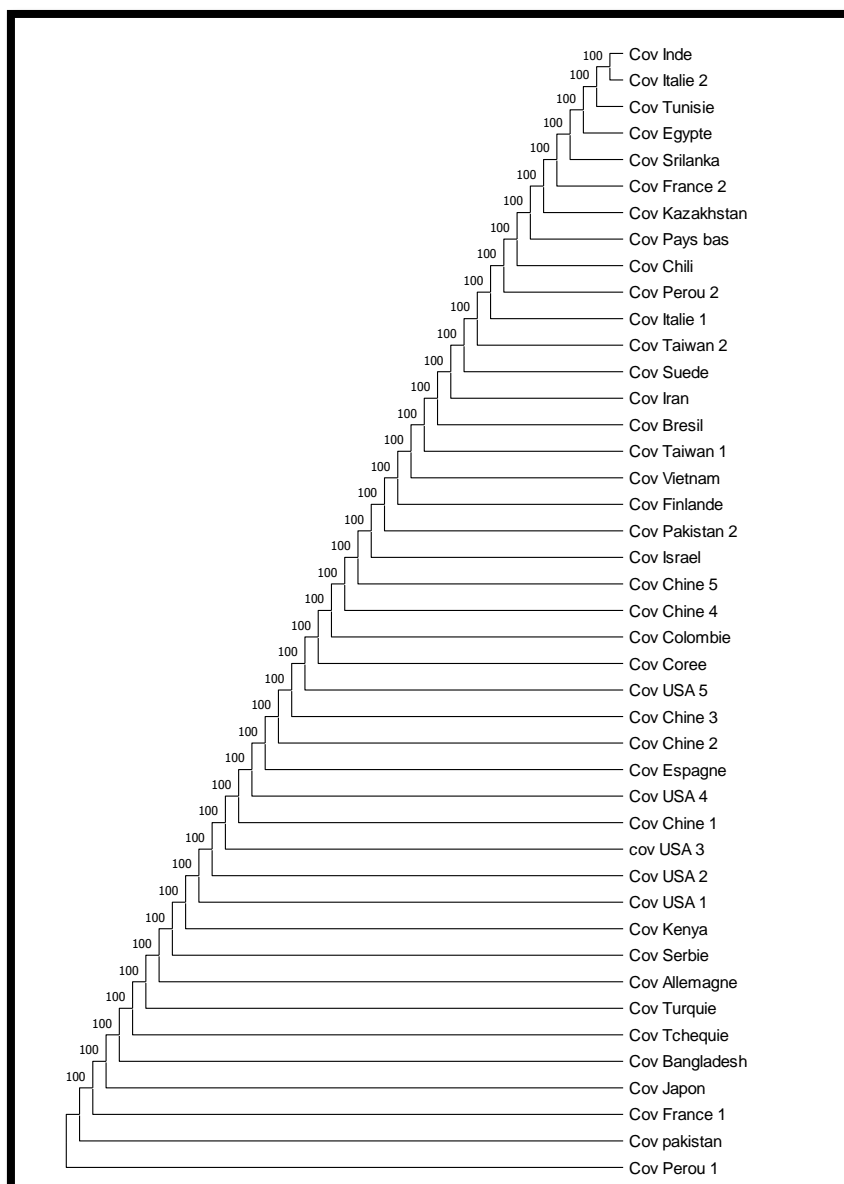


Figure 71 : Arbre phylogénétique de gène E de SARS Cov 2 construit à partir de 43 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

➤ Les résultats de DnBA pour le gène E sont présentés dans le chapitre III figure 37.

3.2. Gene M

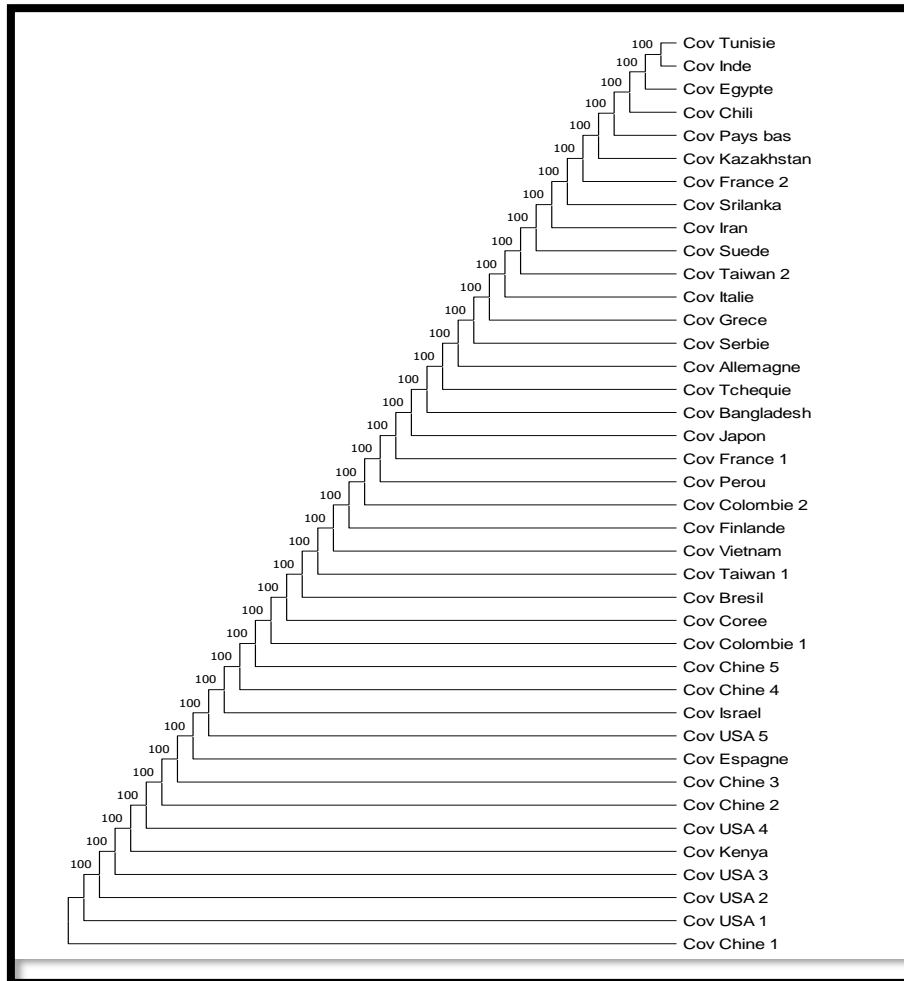


Figure 72 : Arbre phylogénétique de gène M de SARS Cov 2 construit à partir de 40 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

Fichier	Edition	Format	Affichage	?
Chine1:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
USA1:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
USA2:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
USA3:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
USA4:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Chine2:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Chine3:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Espagne:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
USA5:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Israël:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Chine4:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Chine5:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Colombie:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Corée:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Brezil:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Taiwan 1:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Vietnam:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Finlande:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Colombie2:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Perou:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
France1:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Japon:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Bangladesh:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Tcheque:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Allemagne:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Serbie:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Grèce:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Italie:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Taiwan2:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Suede:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Iran:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Sri Lanka:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
France2:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Kazakhstan:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Paysbas:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Chili:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Egypte:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Tunisie:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
Inde:	A: 25.561%	C: 21.824%	G: 20.777%	T: 31.839%
A: 25.561%-25.561% C: 21.824%-21.824% G: 20.777%-20.777% T: 31.839%-31.839%				
min (G-C/C+G) = -0.025 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israël,Chine4,Chine5,Colombie,Corée				
max (G-C/C+G) = -0.025 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israël,Chine4,Chine5,Colombie,Corée				
min (A-T/A+T) = -0.109 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israël,Chine4,Chine5,Colombie,Corée				
max (A-T/A+T) = -0.109 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israël,Chine4,Chine5,Colombie,Corée				

Figure 73 : Analyse des bases azotiques des séquences de gène M par DnBA.

3.3. Gène ORF6

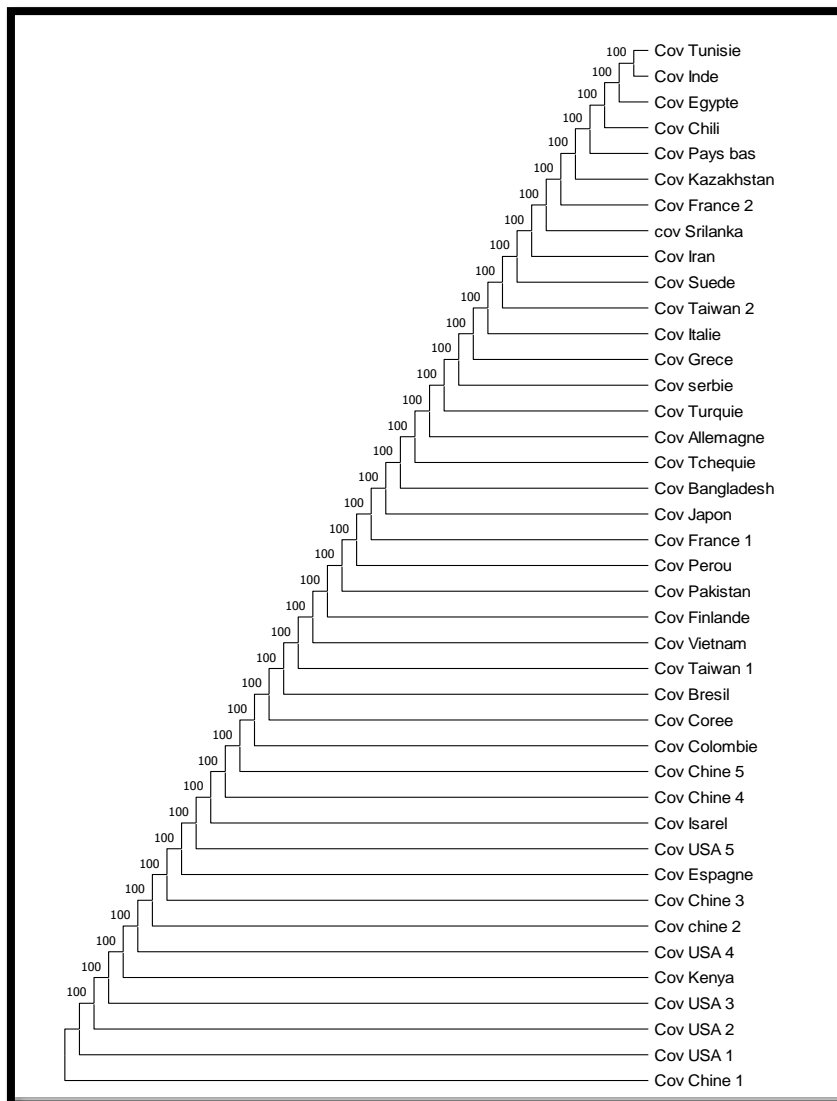


Figure 74: Arbre phylogénétique de gène ORF6 de SARS Cov 2 construit à partir de 41 séquences, en utilisant la méthode UPGMA par le logiciel MEGA.

ORF6.txt - Bloc-notes									
Fichier	Edition	Format	Affichage ?						
Chine1:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
USA1:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
USA2:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
USA3:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Kenie:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
USA4:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Chine2:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Chine3:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Espagne:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
USA5:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Israel:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Chine4:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Chine5:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Colombie:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Coree:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Bresil:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Taiwan1:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Vietnam:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Finlande:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Pakistan:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Perou:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
France1:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Japon:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Bangladesh:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Tcheque:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Allemagne:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Turquie:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Serbie:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Grece:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Italie:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Taiwan2:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Suede:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Iran:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Sri Lanka:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
France2:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Kazakhstan:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Paysbas:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Chili:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Egypte:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Tunisie:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
Inde:	A: 36.559%	C: 13.978%	G: 13.978%	T: 35.484%	(A-T/A+T): 0.000	(G-C/C+G): 0.015	Total: 186pb		
A: 36.559%-36.559% C: 13.978%-13.978% G: 13.978%-13.978% T: 35.484%-35.484%									
min (G-C/C+G) = 0.000 avec 41 pays: Chine1,USA1,USA2,USA3,kenya,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie									
max (G-C/C+G) = 0.000 avec 41 pays: Chine1,USA1,USA2,USA3,kenya,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie									
min (A-T/A+T) = 0.015 avec 41 pays: Chine1,USA1,USA2,USA3,kenya,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie									
max (A-T/A+T) = 0.015 avec 41 pays: Chine1,USA1,USA2,USA3,kenya,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie									

Figure 75: Analyse des bases azotiques des séquences de gène ORF 6 par DnBA.