

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



Faculté de Sciences de la Nature et de la vie

Département De Biologie
Mémoire de fin d'études

En vue de l'obtention du diplôme de Master dans le domaine SNV
Filière Sciences Biologiques
Option : Génétique
Thème

Etude comparative entre des séquences d'ADN de
Coronaviridae chez l'homme et d'autres espèces
animaux.

Présenté par :

M^{elle} Bouguelil Fatma Zohra

M^{elle} Zelelif Ines

Date de la soutenance :

14/09/2021

Devant le jury :

Mme HAMZI W

Mme MOKRANE

Mme Derouiche L

Mr SAID M.

M.C.A USDB 1

M.C.B USDB 1

M.C.A ESSAIA

M.C.A USDB 1

Présidente

Examinatrice

Promotrice

Co-promoteur

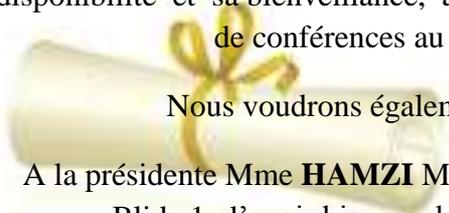
Promotion : 2020 /2021

Remerciements



Nous remercions ALLAH le tout puissant de nous avoir donné le courage, la volonté et la patience ainsi que la force afin de réaliser ce travail.

Nous adressons nos remerciements à toutes les personnes qui nous ont aidé dans la réalisation de ce mémoire. En premier lieu nous tenons, à exprimer nos gratitude et remerciements à notre promotrice **Mme DEROUICHE** Maitre de conférences à l'École Supérieure des Sciences de l'Aliment et Industries Agroalimentaires, Alger, pour sa patience, et sur tout pour sa confiance, ses remarques et ses conseils, sa disponibilité et sa bienveillance, ainsi notre Co-promoteur Mr. **MED SAID** Maitre de conférences au département de biologie, Blida 1.



Nous voudrions également remercier les membres du jury :

A la présidente Mme **HAMZI** Maitre de conférences au département de biologie, Blida 1, d'avoir bien voulu nous faire l'honneur de présider ce jury.

A l'examinatrice Mme **MOKRANE** Maitre de conférences au département de biologie, Blida 1, pour avoir accepté d'évaluer ce travail et pour toutes leurs remarques et critiques.

Nos remerciements s'adressent également à tous nos enseignants, qui nous ont donné les bases de la science.



Je dédie ce mémoire

A ma chère mère pour sa grande aide, sa patience, durant toute ma vie

A mon père

A mes frères et ma sœur

A mes amies Khawla et Assia pour leurs idées et leur soutien morale surtout dans ma maladie que dieux vous protège.

A ma chère binôme Zahra pour son aide et je la remercie d'être avec moi durant ma maladie et durant les moments difficiles que j'ai passé que dieux te protège.





Je dédie ce mémoire

A mon chère père Rabah

A ma chère mère Fatiha

Qui n'ont jamais cessé, de me soutenir et de m'épauler pour que je puisse atteindre mes objectifs.

A mon adorable sœur khadidja et son mari

A ma chère sœur, Hadjer

A mes frères, Mohamed, Ali

A mes chères amies, Chiraz, Zola pour leur support dans les moments difficiles.

Merci d'avoir été là quand j'en avais besoin.

A ma chère binome, Ines pour son soutien morale, sa patience et sa compréhension tout au long de ce mémoire.

A ceux qui ont toujours cru en moi.

Zahra



Listes des Abréviations

HCoV-229E	Human Coronavirus 229E
HCoV-NL63	Human Coronavirus NL63
HCoV-OC43	Human Coronavirus OC43
HCoV-HKU1	Human Coronavirus HKU1
SARS-CoV	Le coronavirus du syndrome respiratoire aigu sévère
MERS-COV	Le coronavirus du syndrome du Moyen-Orient
CSG	Le groupe d'étude sur les Coronaviridae du Comité international de taxonomie
Protéine E	Protéine d'enveloppe
Protéine M	Protéine de membrane
Protéine S	Protéine Spike (pointe)
ORF	Open Reading Frame
Protéine Nsp	Protéine non structurale
Protéine N	Protéine nucléocapside
FCOV	Le coronavirus félin (chat)
BCOV	Le coronavirus de bovins
CCOV	Le coronavirus des Canin (chiens)
TCOV	Le coronavirus de la Turquie (dinde)
PCOV	Le coronavirus de la pintade
IBV	Infectious bronchitis virus
SRAS-COV-2	Le coronavirus du syndrome respiratoire aigu sévère 2 (covid-19)
OMS	Organisation mondiale de la santé
DPP4	Dipeptidyl Peptidase 4
APN	L'aminopeptidase N humaine
ARN	Acide ribonucléique
ACE2	L'enzyme de conversion de l'angiotensine 2
DMV	Vésicules périnucléaires à double membrane
DMS	Sphérules ouvertes à double membrane
ER	Réticulum endoplasmique
RdRp	ARN polymérase ARN-dépendante
RTC	Complexe de réplication et de transcription virale
BLAST	Basic Local Alignment Search Tool.
EMBL	L'Institut Européen de Bioinformatique
DDBJ	Banque Japonaise de données
NCBI	National Center for Biotechnology Information
PHYLIP	Phylogeny Inference Package
MCMC	Méthode Bayesian Markov Chain Monte Carlo
Bio Edit	Editeur d'alignement
MEGA	Moléculaire Génétique évolutive Analyse
UPGMA	Unweighted Pair Group Method of

	Arithmetic Averages
NJ	Neighbor joining
DNASP	DNA Sequence Polymorphism.
PRCV	Le coronavirus respiratoire porcin
TGEV	La gastro-entérite transmissible
PEDV	La diarrhée épidémique porcine
PHEV	L'Encéphalomyélite Hémagglutinante Porcine

Liste des figures

Figure	Titre	Page
1	Classification de mise à jour du HCOV et d'autres coronavirus	4
2	Diagramme schématique d'un virion COV avec l'ensemble minimal de quatre protéines structurales requises pour un assemblage efficace des particules virales infectieuses	5
3	Structure du génome des coronavirus, exemple du SARS-COV	5
4	Les différentes espèces touchées par le coronavirus	6
5	Répartition mondiale des HCOV	9
6	Nombre de cas confirmés de COVID-19 par semaine de notification à la fin de la semaine 17/2021 au 02 mai 2021 en Algérie	10
7	Nombre de cas confirmés de COVID-19 par wilaya du 25 février 2020 au 26 avril 2021	11
8	Le virion et le cycle de vie	12
9	Transmission intra et inter-espèces des coronavirus humains	14
10	Épidémie de syndrome respiratoire aigu sévère ou SARS	16
11	Transmission du coronavirus de type MERS-COV	16
12	Les étapes initiales pour récupérer les séquences de la GenBank	29
13	a : Tableau informatif du Coronaviridae obtenu par la GenBank	29
	b : Quelques souches virales du Coronaviridae non classées obtenues dans la GenBank	30
14	Exemple des séquences nucléotidiques trouvées pour coronavirus non classées de la chauve-souris	30
15	Exemple de séquences nucléotidique du coronavirus de chauve-souris.	31
16	Exemple d'une séquence nucléotidique au format FASTA	32
17	La page principale du programme Fabox	33
18	La conversion des séquences sous forme Fasta par l'utilisation de Fabox	34
19	Alignement trimmer (forme Fasta)	34
20	Barre d'outils de MEGA 5	35
21	Le programme ClustalW utilisé pour l'alignement multiple des séquences du marqueur moléculaire choisi	36
22	La méthode UPGMA utilisée pour la construction de l'arbre phylogénétique	37
23	Les paramètres utilisés pour la reconstruction phylogénétique	37
24	Barre d'outils de DnaSP	38
25	Interface utilisateur de DnaSP	39
26	Barre d'outils du programme DAMBE	40
27	Barre d'outils qui fait les graphes de C/G et A/T	40
28	La page d'accueil de NetWork 10.2	41

29	Ouverture du fichier de données de rdf pour le calcul de MJ NetWork.	42
30	Un fichier sous forme *out a été sauvegardé	43
31	Reconstruction de réseau haplotypique.	43
32	Page d'accueil en ligne de programme FAS to RDF Conversion.	44
33	Arbre phylogénétique de gène S de différentes souches du coronavirus construit à partir de 33 séquences en utilisant la méthode UPGMA par logiciel MEGA5.	50
34	Arbre phylogénétique construit par méthode UPGMA basé sur 35 séquences de gène Rrdp des souches virales du coronavirus en utilisant le logiciel MEGA5.	53
35	L'arbre phylogénétique final construit par la méthode UPGMA basé sur 35 séquences de gène Rdrp des souches du coronavirus en utilisant le logiciel MEGA5.	54
36	Nombre des haplotypes de gène S calculé par DnaSP.	55
37	Nombre des haplotypes de RdRp calculé par DnaSP	56
38	Un exemple des séquences identiques de gène S calculé par dambe	57
39	Les analyses de séquence de coronavirus du ferret par Dambe.	58
40	La variation de GC et AT skews de gène S (séquence du coronavirus de ferret).	58
41	Réseau haplotypique des séquences de gène S réalisé par le logiciel NetWork basé sur les souches virales du coronavirus dans différentes régions géographiques	60
42	Réseau haplotypique des séquences de gène RdRp réalisé par le logiciel NetWork basé sur les souches virales du coronavirus dans différentes régions géographiques	61
43	Réseau haplotypique des séquences de gène S réalisé par le logiciel NetWork basé sur les souches virales du coronavirus dans différentes régions géographiques (méthode 2).	62
44	Réseau haplotypique des séquences de gène RdRp réalisé par NetWork basé sur les souches virales du coronavirus dans différentes régions géographiques (méthode 2).	63

Liste des tableaux

N° de tableau	Titre	Page
I	Les espèces animales touchées par le coronavirus.	7
II	Les sources et les hôtes de l'épidémie du coronavirus humain.	8
III	Séquences nucléotidiques de différentes souches virales du coronavirus extraites de la GenBank.	45
IV	Les variations globales AT et GC skew des gènes S et RdRp	57

Glossaire

- **Haplotype** : Ensemble de gènes situés côte à côte sur un chromosome. Ils sont généralement transmis ensemble à la génération suivante, et sont dits « génétiquement liés ».
- **FASTA** : Algorithme proposé par Pearson basé sur l'identification rapide de segments strictement identiques entre la séquence cible et les séquences de la banque (**Comet, 1998**).
- **Bootstrap** : Méthode statistique pour évaluer la robustesse d'un arbre phylogénétique.
- **UPGMA** : Nom d'un algorithme destiné à la construction d'un arbre phylogénétique. Cette méthode permet la transformation d'une matrice de distances (entre différents organismes, populations, ou séquences de nucléotides) en un arbre enraciné.
- **Maximum de parcimonie** : Approche statistique non-paramétrique très utilisée notamment sur l'inférence phylogénétique.
- **La diversité génétique** : Désigne le degré de variétés des gènes au sein d'une même espèce, correspondant au nombre total de caractéristiques génétique dans la constitution génétique de l'espèce. Elle décrit le niveau de la diversité intraspécifique.
- **Haplogroupe** : Grand groupe d'haplotypes, qui sont des séries d'allèles situés à des sites spécifiques dans un chromosome et qui sont définis par des mutations par polymorphisme nucléotidique singulier.
- **FAS to rdf** : Programme permettant d'obtenir le format rdf directement à base de format fasta obtenu par MEGA
- **Rdf** : Structure de données constituée de nœuds et organisé en graphe.

Résumé

Les virus sont des agents infectieux qui nécessitent un hôte souvent une cellule, parmi eux il y'a les coronavirus qui sont des virus à ARN simple brin sens positif, leur nom est dû à l'apparence d'une couronne solaire. Dans cette étude nous avons inventorié les espèces animales touché par le coronavirus puis nous avons réalisé une comparaison par une analyse phylogénétique entre les souches virales qui ont touché ces animaux avec les souches virales qui ont touché l'homme. Pour la réalisation de ce travail nous avons utilisé des outils de bioinformatiques permettant l'analyse phylogénétique et la détermination des haplotypes, les logiciels utilisés sont, MEGA5, DnaSP, NetWork et DAMBE. Ces analyses sont appliquées sur des séquences d'ADN extraite de la GenBank de deux gènes, RdRp et S. Nous avons classé les souches virales de chaque gène selon l'espèce animale infecté et la zone géographique où il a été retrouvé.

Nous avons travaillé sur 33 séquences nucléotidiques du gène S de différentes souches telles que SARS-CoV1 et MERS-CoV infectant l'homme, Bat-CoV infectant les chauves-souris, PEDV et PDCOV infectant le porc, à travers plusieurs pays comme la Chine, l'Arabie Saoudite, les Etats-Unis, le Japon, l'Espagne. Nous avons trouvé qu'il y a une accumulation des mutations au sein du gène S entre les différentes souches et que certains changements de nucléotides sur ce gène de pointe peuvent provoquer une transmission du virus de l'animal à l'homme. L'analyse du gène RdRp a été basé sur 32 séquences nucléotidiques de différentes souches virales ; les résultats obtenus prouvent la présence des souches Bat-Cov infectant les chauves-souris dans presque tous les continents du monde, RtCov infectant les rongeurs en Asie, HCov/229 qui touche l'homme au Brésil et en Slovénie, PEDV et PHEV du porc en Italie, BCov touchant les bovins en Slovénie, FRSCV des furets au Pérou, et ECov des hérissons en Chine. Sachant que le coronavirus est considéré comme l'un des plus anciens virus il n'est pas surprenant que la RdRp virale est une polymérase évolutive primitive et les mutations de ce gène conduisent à l'apparition des différentes souches de virus affectant différentes espèces dans des zones géographiques variables.

Les résultats obtenus des deux gènes S et RdRp nous ont démontré que la diversité génétique entre les souches de coronavirus est très élevée que ce soit pour les souches qui infectent les animaux ou celles qui infectent l'homme dans différents pays du monde

Mots clé : Coronaviridae, gène, phylogénie, bioinformatique, ADN.

Summary

Viruses are infectious agents that require a host often a cell, among them there are coronaviruses which are single strand RNA viruses positive sense, their name is due to the appearance of a solar corona. In this study we inventoried the animal species affected by the coronavirus and then carried out a phylogenetic analysis comparison between the viral strains that affected these animals and the viral strains that affected humans. For the realization of this work we used bioinformatics tools allowing phylogenetic analysis and the determination of haplotypes, the software used are, MEGA5, DnaSP, NetWork and DAMBE. These analyses are applied to DNA sequences extracted from GenBank from two genes, RdRp and S. We classified the viral strains of each gene according to the infected animal species and the geographical area where it was found.

We have worked on 33 nucleotide sequences of the S gene from different strains such as SARS-CoV1 and MERS-CoV infecting humans, Bat-CoV infecting bats, PEDV and PDCOV infecting pigs, across several countries such as China, Saudi Arabia, the UnitedUnited States, Japan, Spain. We found that there is an accumulation of mutations within the S gene between the different strains and that certain nucleotide changes on this advanced gene can cause virus transmission from animal to human. Analysis of the RdRp gene was based on 32 nucleotide sequences of different viral strains; the results prove the presence of Bat-Cov strains infecting bats in almost every continent of the world, RtCov infecting rodents in Asia, HCov/229 affecting humans in Brazil and Slovenia, PEDV and PHEV of pork in Italy, BCov affecting cattle in Slovenia, FRSCV of ferrets in Peru, and ECov of hedgehogs in China. Knowing that the coronavirus is considered one of the oldest viruses it is not surprising that viral RdRp is a primitive evolutionary polymerase and mutations of this gene lead to the appearance of different strains of viruses affecting different species in variable geographic areas.

The results obtained from the two genes S and RdRp have shown us that the genetic diversity between the strains of coronavirus is very high whether it is for strains that infect animals or those that infect humans in different countries of the world

Keywords: Coronaviridae, gene, phylogeny, bioinformatics, ADN.

ملخص

الفيروسات هي العوامل المعدية التي تتطلب مضيف في كثير من الأحيان خلية، من بين هذه الفيروسات هناك coronaviruses التي هي فيروسات RNA أحادية الشريط حاسة إيجابية، اسمها يرجع إلى ظهور كورونا الشمسية. في هذه الدراسة قمنا بجرد الأنواع الحيوانية المتأثرة بفيروس كورونا ومن ثم قمنا بإجراء مقارنة بين السلالات الفيروسية التي أثرت على هذه الحيوانات والسلالات الفيروسية التي أثرت على البشر. ولتحقيق هذا العمل استخدمنا أدوات المعلوماتية الحيوية التي تسمح بالتحليل الفيلوجيني وتحديد النماذج، MEGA5 البرمجيات المستخدمة هي NA SP و NetWork و DAMBE وتطبق هذه التحاليل على تسلسلات الحمض النووي المستخرجة من جينك من جينين هما RdRp و S. قمنا بتصنيف السلالات الفيروسية لكل جين حسب الأنواع الحيوانية المصابة والمنطقة الجغرافية التي تم العثور عليها فيها.

وقد عملنا على 33 سلسلة نوكلويدات من الجين S من سلالات مختلفة مثل SARS-CoV1 و MERS-CoV التي تصيب البشر، وخفافيش Bat-Cov التي تصيب الخنازير، PEDV، PDCOV التي تصيب الخنازير، عبر عدة بلدان مثل الصين والمملكة العربية السعودية والولايات المتحدة واليابان وإسبانيا. وجدنا أن هناك تراكم للطفرات داخل جين S بين السلالات المختلفة وأن تغيرات نووية معينة على هذا الجين المتقدم يمكن أن تسبب انتقال الفيروس من الحيوان إلى الإنسان. واستند تحليل جين RDRP إلى 32 تسلسل نووي من سلالات فيروسية مختلفة؛ وتثبت النتائج وجود سلالات من "Bat CoV" تصيب الخفافيش في كل قارة من قارات العالم تقريبا، HCov/229 تصيب القوارض في آسيا، وتؤثر على البشر في البرازيل وسلوفينيا، PEDV تؤثر على الخنزير في إيطاليا، و BCov تؤثر على الماشية في سلوفينيا، FRSCV تؤثر على النمس في بيرو، و ECov تؤثر على القنفذ في الصين. مع العلم أن فيروس كورونا فيروس يعتبر واحد من أقدم الفيروسات، فإنه ليس من المستغرب أن الفيروسي RdRp هو بوليميراز تطوري بدائي والطفرات من هذا الجين يؤدي إلى ظهور سلالات مختلفة من الفيروسات التي تؤثر على الأنواع المختلفة في المناطق الجغرافية المتغيرة.

النتائج التي تم الحصول عليها من الجينين S و RdRp أظهرت لنا أن التنوع الجيني بين سلالات فيروس coronavirus مرتفع جدا سواء كان للسلالات التي تصيب الحيوانات أو تلك التي تصيب البشر في مختلف دول العالم

الكلمات الرئيسية: Coronaviridae: جين، phylogeny، المعلوماتية الحيوية، ADN

Sommaire

INTRODUCTION.....	1
CHAPITRE I : RAPPELS BIBLIOGRAPHIQUES	
I.1. Présentation de la famille Coronaviridea	4
I.2.Structure de coronavirus.	9
I.3.Les espèces touchées par le coronavirus.....	5
I.4.Le coronavirus chez l'homme	7
I.4.1. Propagation des souches de coronavirus humain dans le monde	9
I.4.1.1. Propagation de covid-19 en Algérie.....	10
I.4.2. Le cycle de vie de coronavirus humain et réplication.....	11
I.4.2.1. Le cycle de vie.....	11
I.4.2.2. Réplication	13
I.4.3. Transmission virale de coronavirus humain.....	14
I.4.3.1. Transmission virale de SRAS-COV.....	15
I.4.3.2. Transmission virale de MERS-COV.....	16
I.4.3.3. Transmission virale de SARS-COV-2.....	17
I.5.Identification phylogénétique.....	17
I.5.1 Définition de la phylogénie.....	17
I.5.2. Phylogénie moléculaire.....	18
I.5.3. L'arbre phylogénétique.....	19
I.6. La phylogénie et l'informatique.....	20
I.6.1. Les bases de données.....	20
I.6.2. Les banques généralistes.....	20
I.6.2.1. Banques de séquences nucléotidiques.....	20
I.6.2.2. Banques de séquences protéiques	21
I.6.3. Les banques spécialistes.....	23
I.7. Les logiciels utilisé dans la phylogénie.....	24

I.7.1.1. Les méthodes de reconstruction phylogénétique et les logiciels utilisés.....	24
I.7.1.1.1. La méthode de parcimonie.....	24
I.7.1.1.2. La méthode Bayésien inference	25
I.7.1.1.3. La méthode de Maximum de vraisemblance.....	25
I.7.1.2. Les logiciels de Nettoyage de l’alignement.....	26
I.7.1.3. Les logiciels pour l’alignement de séquence.....	26
I.7.1.4. Les logiciels de traitement de séquence.....	26
I.7.1.5. Logiciels de visualisation et reconstruction d’arbres phylogénétiques.....	27
 CHAPITRE II : MATERIEL ET METHODES	
II.1. Analyses moléculaires et phylogénétiques	28
II.2. Extraction des données	28
II.2.1. Définition de la GenBank	28
II.2.1.1. Traitement des séquences de la GenBank.....	32
II.2.1.1.1. MEGA version 5.....	32
II.2.1.1.2. FABOX.....	33
II.2.1.1.3. DNASP.....	38
II.2.1.1.4. Dambe.....	39
II.2.1.1.5. Network	40
 CHAPITRE III : RESULTATS ET DISCUSSION	
III.1. Résultats	45
III.1.1.Genbank.....	45
III.1.2. Construction des arbres phylogénétiques par MEGA.....	48
III.1.2.1. Arbre phylogénétique de gène de pointe (S).....	50
III.1.2.2. Arbre phylogénétique de gène RdRp.....	55
III.1.3. DnaSP.....	55
III.1.4. Dambe.....	57
III.1.5. Network.....	60
III.2. Discussion	65

III.2.1 La phylogénie de coronavirus.....	65
III.2.1.1 Les relations phylogénétiques au sein d'arbre S.....	65
III.2.1.1.2 Les relations phylogénétiques au sein d'arbre Rdrp.....	68
III.2.2. Comparaison générale entre les logiciels utilisés.....	71
CONCLUSION.....	72
RÉFÉRENCES BIBLIOGRAPHIQUES.....	74
BIBLIO-WEB	
ANNEXE	

Introduction

La famille des Coronaviridae est une famille comporte un groupe monogénique de 11 virus infectant les vertébrés (**Siddell et Coll, 1983**).

Selon le Comité international de taxonomie des virus, les coronavirus sont classés sous l'ordre des *Nidovirales*, famille des *Coronaviridae*, sous-famille des *Coronavirinae*. Sur la base de preuves sérologiques précoces et génomiques ultérieures, *Coronavirinae* est divisé en quatre genres: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* et *Deltacoronavirus* (**Fun, 2019**). Ce virus est considéré comme l'agent causal de différentes maladies respiratoires et digestifs.

Avant 2003, il n'y avait que 10 coronavirus avec des génomes complets disponibles. Après l'épidémie de SRAS, jusqu'en décembre 2008, il y a eu un ajout de 16 coronavirus avec des génomes complets séquencés. Ceux-ci comprennent deux coronavirus humains (coronavirus humain NL63 et coronavirus humain HKU1), 10 autres coronavirus de mammifères [coronavirus du SRAS de chauve-souris, coronavirus de chauve-souris (bat-CoV) HKU2, bat-CoV HKU4, bat-CoV HKU5, bat-CoV HKU8, bat-CoV HKU9, bat-CoV 512/2005, bat-CoV 1A, coronavirus équin et coronavirus de béluga] et quatre coronavirus aviaires (coronavirus de dinde, coronavirus bulbul HKU11, coronavirus de muguet HKU12 et coronavirus de Munia HKU13) (**Patrick et al., 2009**).

La phylogénétique est une composante importante de l'approche de la biologie des systèmes (**Kuhl et al., 2019**), elle a comme but l'étude des liens de parenté entre les espèces et leur histoire d'évolution et la diversification des organismes (**Cornwell et al., 2017**).

La bioinformatique est devenue une partie intégrante de la recherche et du développement en sciences dans ces dernières années et a désormais un rôle essentiel à la fois dans le décryptage des données génomiques, transcriptomiques et protéomiques générées par les technologies expérimentales à haut débit et dans l'organisation des informations recueillies en biologie traditionnelle. Des méthodes basées sur des séquences d'analyse de gènes ou de protéines individuels ont été élaborées et étendues, et des méthodes ont été développées pour analyser un grand nombre de gènes ou de protéines (**Kahenisa et al., 2003**) sur le plan moléculaire et phylogénétique.

Notre thème est fondé sur une étude phylogénétique du coronavirus qui est le virus le plus célèbre et le plus diffus dans le monde entier actuellement. Ce qui est effectivement difficile de déterminer l'agent causal de la pandémie issu de ce virus et ainsi de comprendre son comportement sachant que le coronavirus est un virus exposé à un taux élevé de mutations. Nous avons commencé notre travail par des connaissances fondamentales sur la taxonomie de coronavirus en général et ces souches virales en particulier et leur diffusion dans différentes zones géographiques et aussi par à part les espèces touchées concernant les deux gènes : ARN-polymérase ARN-dépendante et le gène de protéine de pointe. Ensuite nous avons appliqué les éléments de base de la phylogénie qui reposent sur la construction de l'arbre phylogénétique localisant l'évolution de coronavirus et la comparaison entre les souches en utilisant plusieurs logiciels de bioinformatique qui nous a permis et facilité à la suite d'interpréter et discuter notre résultat obtenu.

Enfin, nous avons finalisé notre travail par une conclusion générale où nous avons libéré les résultats essentiels ainsi que les prévisions ouvertes de cette étude.

I.1. Présentation de la famille Coronaviridea

La famille des *Coronaviridae* comprend un groupe monogérique de 11 virus infectant les vertébrés (Siddell et Coll, 1983).

Selon le Comité international de taxonomie des virus, les coronavirus sont classés sous l'ordre des *Nidovirales*, famille des *Coronaviridae*, sous-famille des *Coronavirinae*. Sur la base de preuves sérologiques précoces et génomiques ultérieures, *Coronavirinae* est divisé en quatre genres: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* et *Deltacoronavirus*. Quatre lignées distinctes (A, B, C et D) ont été attribuées dans le genre *Betacoronavirus*. Parmi les six coronavirus humains connus (HCoV) : HCoV-229E et HCoV-NL63 appartiennent à *Alphacoronavirus*, tandis que HCoV-OC43 et HCoV-HKU1 appartiennent à la lignée A, SARS-CoV à la lignée B et MERS-CoV à la lignée C du genre *Betacoronavirus*. Les autres coronavirus tels que Avion coronavirus appartiennent au genre *Gammacoronavirus* (Fung et al., 2019). Nous représentons ça dans la figure 1.

Après l'épidémie actuelle d'une maladie respiratoire aiguë associée au coronavirus, le groupe d'étude sur les *Coronaviridae* (CSG) du Comité international de taxonomie des virus est chargé de développer la classification des virus et la nomenclature des taxons de la famille des *Coronaviridae*, a évalué le placement du pathogène humain, provisoirement nommé 2019-nCoV, dans le *Coronaviridae*, le CSG reconnaît que ce virus forme un clade sœur du prototype des coronavirus du syndrome respiratoire aigu sévère humain et de la chauve-souris (*SRAS-CoV*) de l'espèce *Coronavirus* lié au syndrome respiratoire aigu sévère, et le désigne comme *SARS-CoV-2* (CSG du Comité international de taxonomie des virus, 2020).

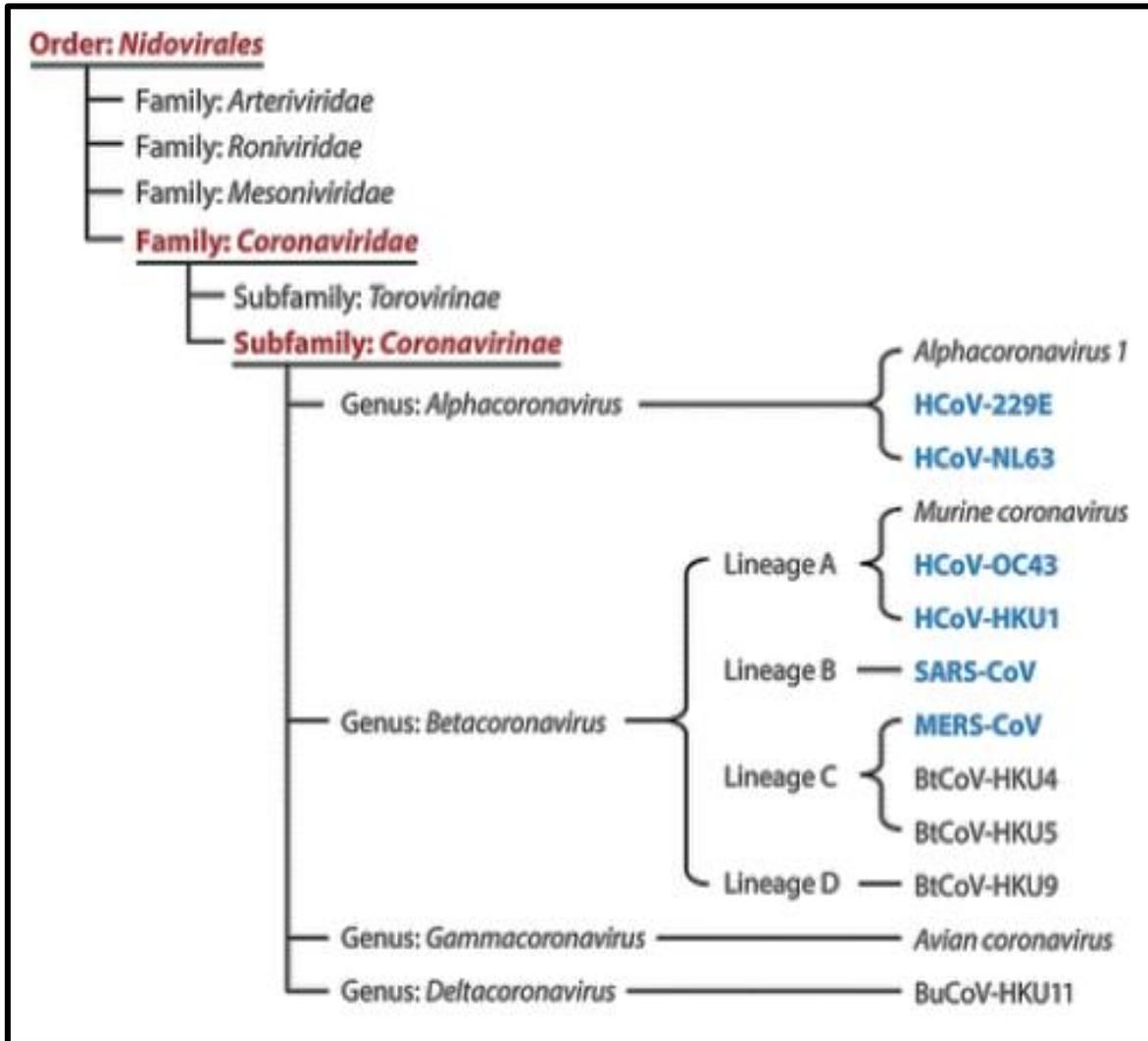


Figure 1 : Classification de mis à jour du HCOV et d'autre coronavirus (Fung, 2019).

I.2. Structure de coronavirus

Les coronavirus doivent leur nom à leur aspect en microscopie électronique, avec des spicules formant une couronne autour de la particule virale (Segondy, 2020).

La particule virale a une taille de l'ordre de 100 à 150 nm. Elle est constituée par une nucléocapside formée par la protéine N liée à l'ARN génomique. Cette protéine N a également des interactions avec les protéines d'enveloppe. L'enveloppe est constituée par un double feuillet lipidique au sein duquel sont insérées les trois protéines d'enveloppe E, M et S (Figure 2). A la partie 5' du génome (Figure 3) les cadres de lecture Orf1a et ORF1b codent 16 protéines non structurales (ns1 à ns16). A la partie 3' se trouvent les cadres de lecture codant les protéines structurales S (spike) E (enveloppe), M (membrane) et N (nucléocapside) (Segondy, 2020).

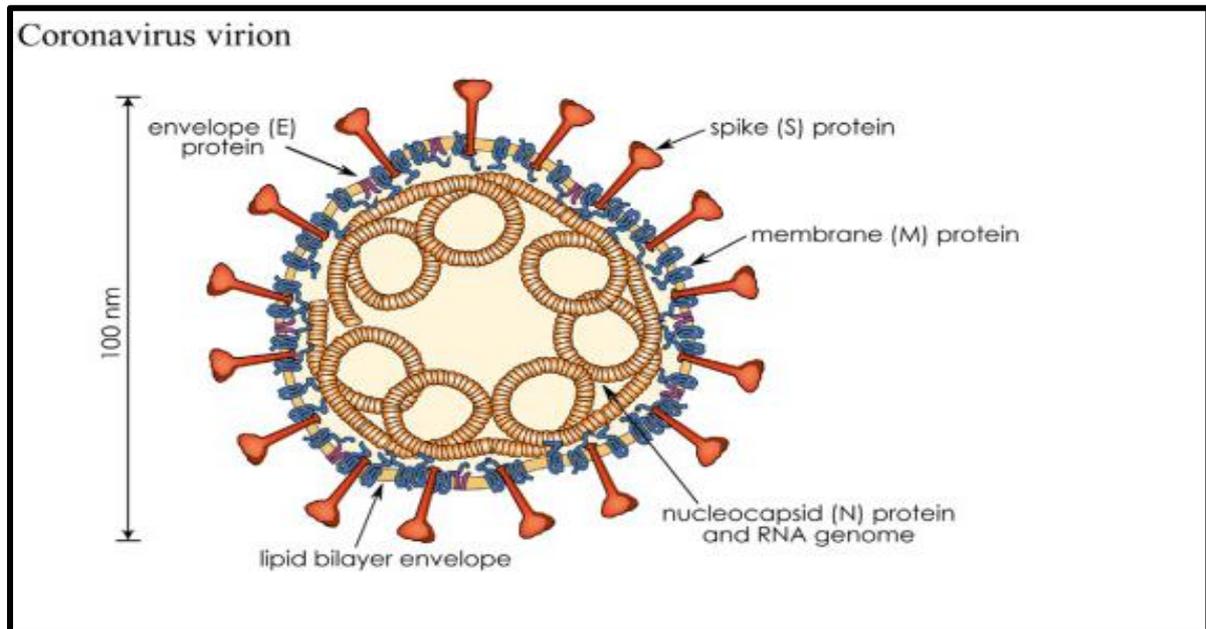


Figure 2 : Diagramme schématisique d'un virion COV avec l'ensemble minimal de quatre protéines structurales requises pour un assemblage efficace des particules virales infectieuses (Deng et al., 2021).

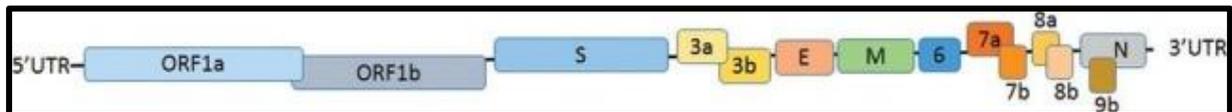


Figure 3 : Structure du génome des coronavirus, exemple du SARS-COV (Segondy, 2020).

I.3. Les espèces touchées par le coronavirus

Sachant que y a plusieurs espèces touchées par le coronavirus (les mammifères, les aviaires.) (Figure 4), nous citons :

- Le coronavirus félin (*FCOV*) (Jaimes et al., 2017).
- Le bovins coronavirus (*BCOV*) (Vlasova et al., 2021).
- Le coronavirus équin (*ECoV*) (Pasterela et al., 2018).
- Des coronavirus ont été isolés de mammifères marins, dont le béluga (*Delphinapterus leucas*) et le grand dauphin (*Tursiops aduncus*) ainsi que d'espèces de cyprinidés d'eau douce comme la dorade blanche (*Blicca bjoerkna*) et le vairon à grosse tête (*Pimephales promelas*) (Schutze, 2016).

- Le coronavirus des chiens (*CCOV*) (Takano et al., 2016).
- Coronavirus de la Turquie (*TCOV*) : le coronavirus de la dinde (Chen et al., 2015).
- Les coronavirus de la pintade (*PCOV*) (Liais et al., 2014).
- Le coronavirus du syndrome respiratoire aigu sévère (*SRAS-COV*) et Le coronavirus du syndrome du Moyen-Orient (*MERS-COV*) (Segondy, 2020).

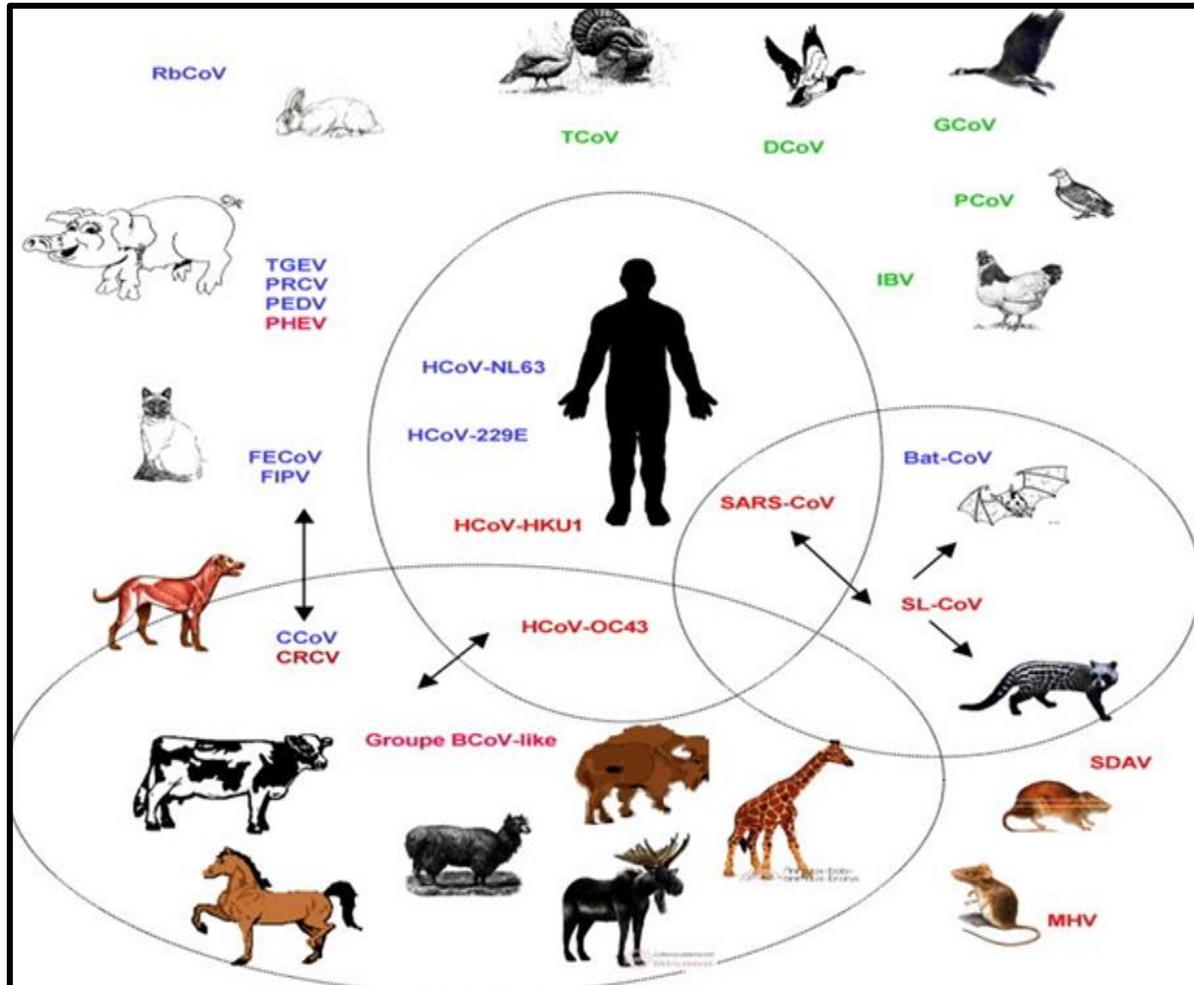


Figure 4 : Les différentes espèces touchées par le coronavirus (Pathol, 2009).

- Nous avons construit le **tableau I** suivant qui résume les souches virales telles que *BCOV*, *CCOV*, *IBV* et leurs effets sur l'organe d'animal touché avec leur voie de transmission et leurs taux de mortalité à travers les travaux d'**Alluwaimi (2020)** sur les coronavirus des animaux et des oiseaux.

Tableau I: Les espèces animales touchées par le coronavirus (**original**).

Souche virale	Espèce touché	L'infection	Résultat d'infection (maladie)	Voie de transmission	Mortalité
BCOV	Bovine	Infecte les voies respiratoires supérieures et inférieures et l'intestin.	Une diarrhée néonatale des veaux, une dysenterie hivernale chez les bovins adultes et des troubles des voies respiratoires chez les bovins de tous âges.	Transmissible par voie zoonotique entre les espèces, car des virus de type BCoV ont été détectés chez des ruminants sauvages et des humains.	Les taux de mortalité élevés lorsque l'infection est associée à des facteurs, principalement le stress ; transport la fièvre, une pneumonie.
CCOV	Chien	Infecte les cellules épithéliales de l'intestin grêle parfois les voies respiratoires supérieures.	Une gastro-entérite et entraînant une diarrhée mortelle.	Transmissible par voie orale	Une faible mortalité.
IBV	Oiseaux	Infecte les voies respiratoires supérieures et les surfaces épithéliales du tube digestif, reins.	Provoquant une baisse spectaculaire de la production d'œuf.	Il se propage par aérosols.	Les pertes élevées dans la production de volaille infligées par l'infection à l'IBV.

I.4. Le coronavirus chez l'homme

Parmi les souches infectant l'homme nous avons le coronavirus du syndrome respiratoire aigu sévère (*SRAS-COV*) ; le coronavirus du syndrome du Moyen-Orient (*MERS-COV*) et le nouveau *SRAS-COV-2* responsable de la maladie à coronavirus 2019 (covid-19) qui cause une baisse sévère les infections des voies respiratoires avec des atteintes extra-pulmonaires sont

dues au *Bétacoronavirus*. Les autres coronavirus humains (*HCoV-NL63*, *HCoV-229E*, *HCoV-OC43* et *HKU1*) qui induisent des maladies respiratoires légères (Cui et al., 2019). Les sept coronavirus responsables d'infections humaines se répartissent dans les genres *Alphacoronavirus*, *Betacoronavirus*.

Les coronavirus, y compris les coronavirus humains endémiques, ont une origine animale. C'est à partir des chauves-souris ou de rongeurs que ces virus se sont diversifiés et ont pu se répandre chez d'autres espèces y compris l'homme (Tableau II) (Corman et al., 2018).

Tableau II : Les sources et les hôtes de l'épidémie de coronavirus humain (Corman et al., 2018).

Coronavirus	Hôtes naturels	Hôtes intermédiaires.
<i>HCoV-NL63</i>	Chauves-souris	Indéterminé
<i>HCoV-229E</i>	Chauves-souris	Dromadaire
<i>HCoV-OC43</i>	Rongeurs	Bovins
<i>HCoV-HKU1</i>	Rongeurs	Indéterminé
<i>SARS-COV-1</i>	Chauves-souris	Civette palmiste masqué
<i>SARS-COV-2</i>	Chauves-souris	Pangolin
<i>Mers-COV</i>	Chauves-souris	Dromadaires

I.4.1. Propagation des souches de coronavirus humain dans le monde

La propagation est résumée dans la figure 5, en général 229E, OC43 et NL63 sont distribués dans le monde (**Figure 5A**), et ont tendance à être transmis principalement pendant la saison hivernale dans les pays à climat tempéré (**Handly, 1972**) tandis que NL63 a montré un pic d'activité printemps-été d'après une étude à Hong Kong (**Chiu, 2005**).

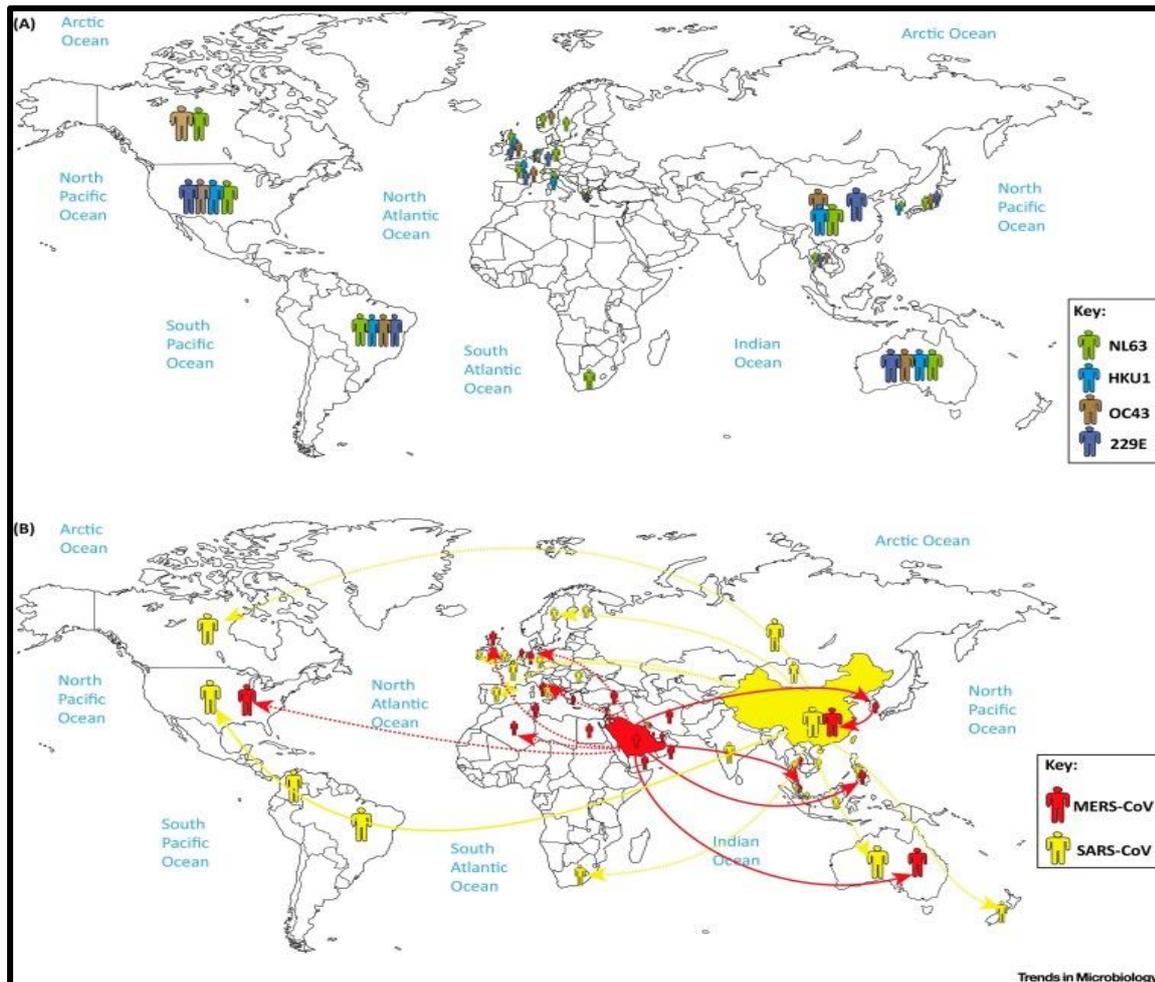


Figure 5 : Répartition mondiale des HCoV (Su et al., 2016).

La **Figure 5A** : le vert, le bleu, le marron et le violet représentent la distribution mondiale des coronavirus humains NL63, HKU1, OC43 et 229E, respectivement

La **Figure 5B** : le rouge et le jaune représentent la distribution mondiale du MERS-CoV et du SARS-CoV, respectivement.

I.4.1.1. Propagation de covid-19 en Algérie

Le 1^{er} cas, un ressortissant italien, a été notifié le 25 février 2020 dans une base de vie à Hassi Messaoud dans la wilaya de Ouargla. A partir du 02 mars 2020 un foyer a été détecté dans la wilaya de Blida pour donner suite à une alerte lancée par la France après la confirmation au COVID-19 de deux citoyens Algériens résidant en France ayant séjourné en Algérie (organisation mondiale de la santé, OMS 2020).

Para port la situation de Covid-19 en Algérie nous chosions la période entre S09/2020 jusqu’ à la S16/2021 qui est présenté dans la figure 6 suivante :

Le nombre le plus élevé de cas notifiés en une semaine depuis le début de la pandémie a été enregistré durant la semaine N°48 (23-29 novembre 2020) avec 7 359 cas. Après une diminution en trois étapes (diminution très significative pendant 6 semaines jusqu’à atteindre 1736 cas, plateau pendant 5 semaines autour de 1700 cas et diminution modérée pendant 6 semaines jusqu’à atteindre 679 cas), le nombre hebdomadaire de cas est reparti à la hausse depuis la semaine N°12/2021 (22-28 mars 2021) sur les 5 dernières semaines (OMS, 2021).

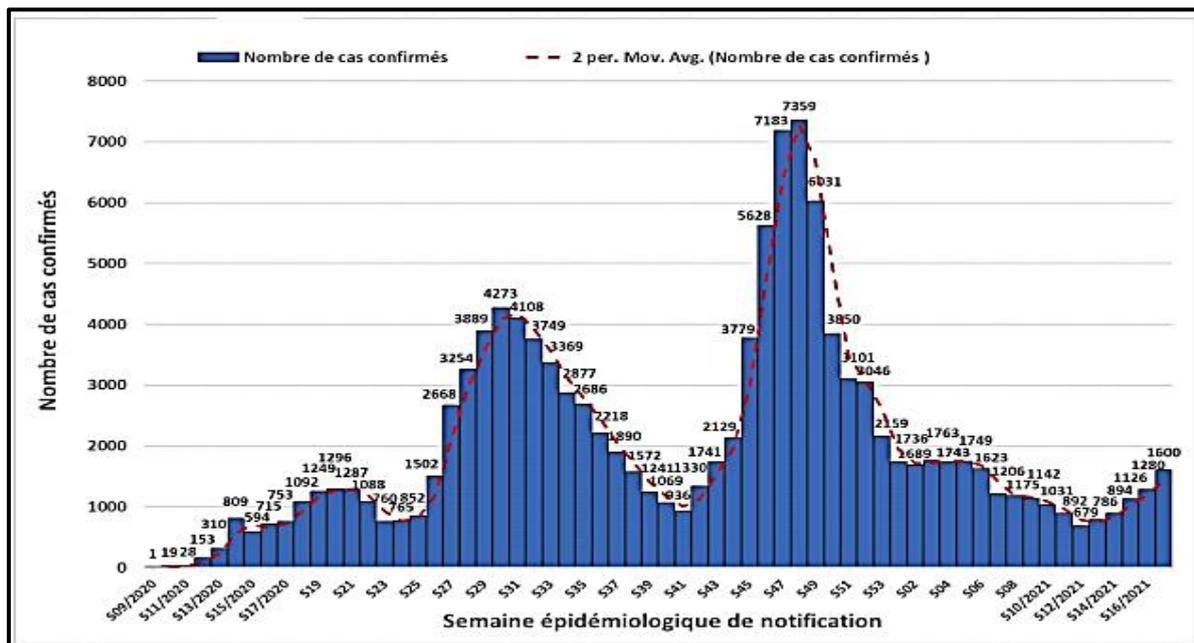


Figure 6 : Nombre de cas confirmés de COVID-19 par semaine de notification à la fin de la semaine 17/2021 au 02 mai 2021 en Algérie (OMS, 2021).

- En date du 26 avril 2021, les wilayas d’Alger, Oran et Blida regroupaient toujours le plus grand nombre de cas notifiés avec 19 032, 12 564 et 8974 cas respectivement (Figure 7), soit 33% des cas déclarés dans le pays. En termes d’incidence, ces wilayas restent toujours les 03 wilayas ayant les taux d’incidence les plus élevés, suivies des

wilayas de Jijel et de Bejaia (incidence supérieure à 450 cas pour 100 000 habitants dans ces 05 wilayas). Sur 14 jours (13 au 26 avril), quatre wilayas ont eu un taux d'accroissement supérieur à 5% : la wilaya de Tindouf (13,5%) avec 33 cas, la wilaya de M'Sila (8,1%), la wilaya de Laghouat (7,3%) et la wilaya d'Ouargla (5,7%). Une seule wilaya n'a pas présenté de cas pendant ces 14 jours : la wilaya de Naâma.

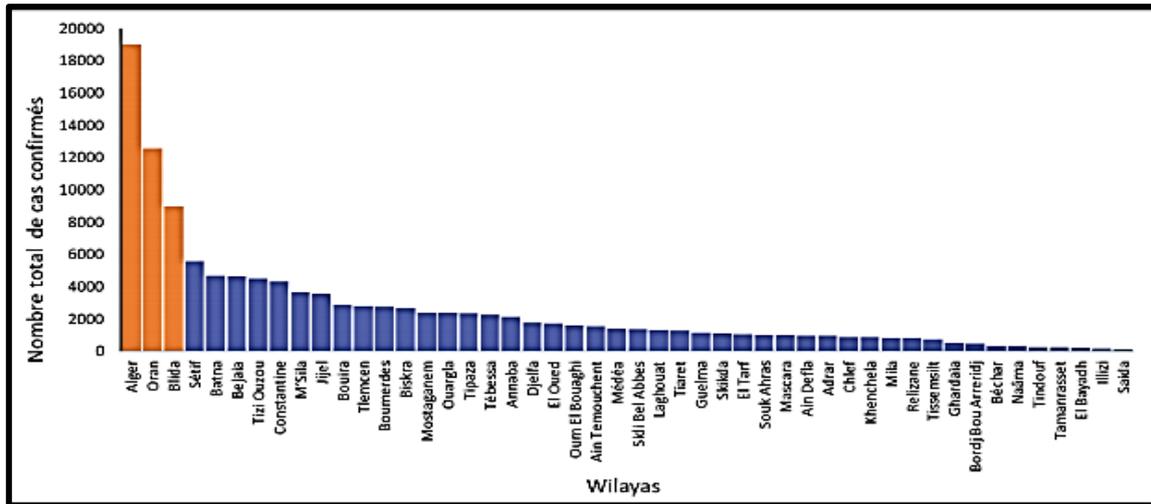


Figure 7 : Nombre de cas confirmés de COVID-19 par wilaya du 25 février 2020 au 26 avril 2021 (OMS, 2021).

I.4.2. Le cycle de vie de coronavirus humain et répliation

I.4.2.1. Le cycle de vie

Les étapes initiales de l'infection à coronavirus impliquent la liaison spécifique de la protéine du pic de coronavirus (S) aux récepteurs d'entrée cellulaire, qui ont été identifiés pour plusieurs coronavirus et comprennent l'aminopeptidase N humaine (*APN ; HCoV-229E*), l'enzyme de conversion de l'angiotensine 2 (*ACE2 ; HCoV-NL63, SARS-CoV et SARS-CoV-2*) et dipeptidyl peptidase 4 (*DPP4 ; MERS-CoV*), l'expression et la distribution tissulaire des récepteurs d'entrée influencent par conséquent le tropisme viral et la pathogénicité. Au cours du cycle de vie intracellulaire (Philip et al., 2020) (Figure 8).

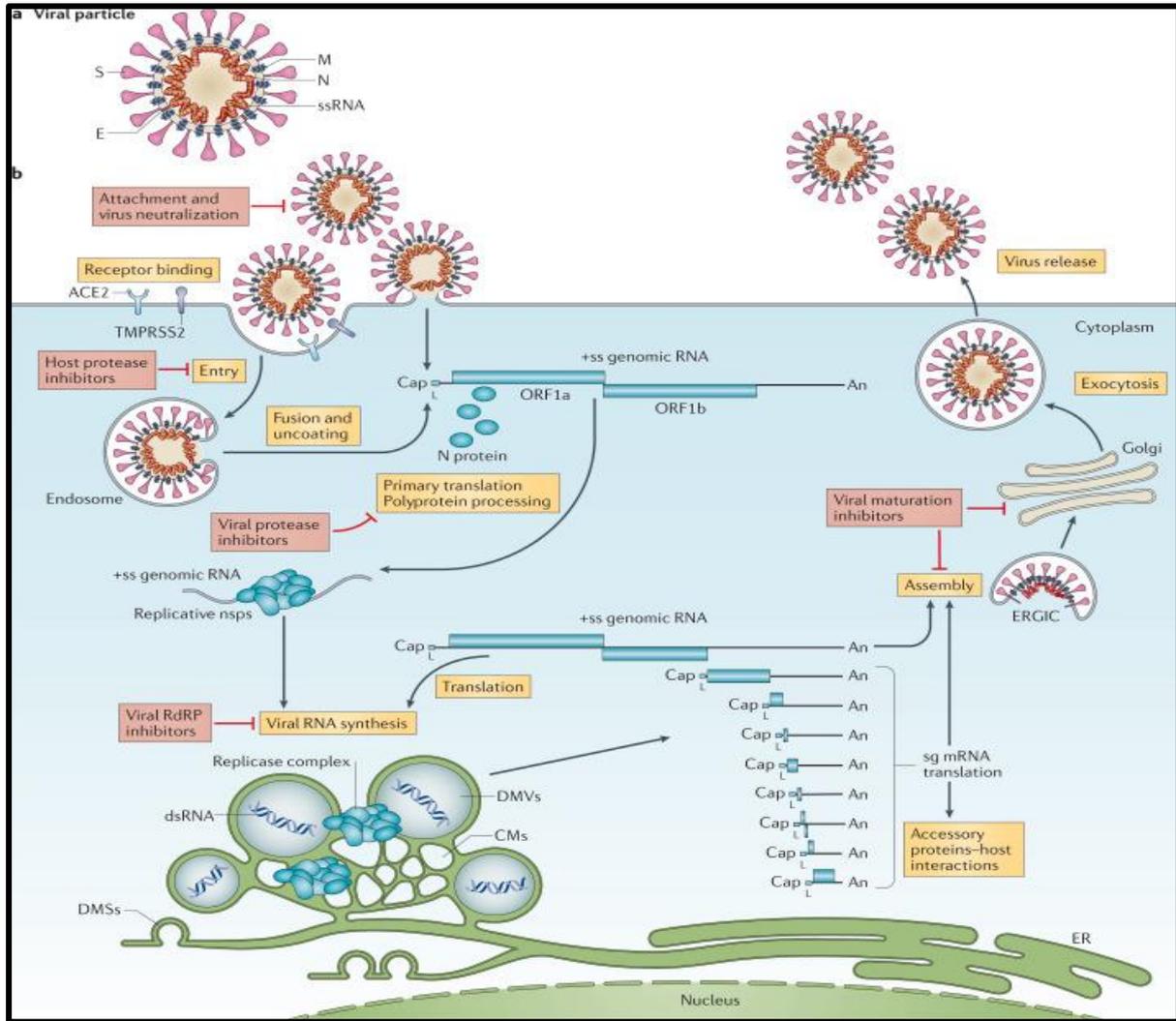


Figure 8 : Le virion et le cycle de vie (Philip, 2020).

Figure 8A : le virion du coronavirus est constitué de protéines structurales, à savoir le pic (S), l'enveloppe (E), la membrane (M), la nucléocapside (N) et, pour certains *Bétacoronavirus*, l'hémagglutinine-estérase (non représentée). Le génome d'ARN simple brin de sens positif (+ ARNss) est encapsidé par N, tandis que M et E assurent son incorporation dans la particule virale lors du processus d'assemblage. Les trimères S dépassent de l'enveloppe virale dérivée de l'hôte et fournissent une spécificité pour les récepteurs d'entrée cellulaires.

Figure 8B : Les particules de coronavirus se lient aux facteurs d'attachement cellulaire et aux interactions S spécifiques avec les récepteurs cellulaires (tels que l'enzyme de conversion de l'angiotensine 2 (ACE2), ainsi que des facteurs hôtes (tels que la sérine protéase de surface cellulaire *TMPRSS2*), favorisent l'absorption virale et la fusion au niveau de membrane cellulaire ou endosomale. Après l'entrée, la libération et le décapage de l'ARN génomique entrant le soumettent à la traduction immédiate de deux grands cadres de lecture ouverts,

ORF1a et ORF1b. Les poly protéines pp1a et pp1ab résultantes sont transformées par co-traduction et post-traduction en les protéines non structurales individuelles (NSPS) qui forment le complexe de réplication et de transcription virale. En concordance avec l'expression du nsps, la biogenèse des organites de réplication virale constituées de vésicules périnucléaires à double membrane (DMV) caractéristiques, de membranes alambiquées (CM) et de petites sphérules ouvertes à double membrane (DMS) créent un microenvironnement protecteur pour la réplication de l'ARN génomique viral et transcription d'ARNm sous-génomiques (ARNm sg) comprenant l'ensemble imbriqué caractéristique d'ARNm de coronavirus. Les protéines structurales traduites se transloquent dans les membranes du réticulum endoplasmique (ER) et transitent par le compartiment intermédiaire ER-à-Golgi (ERGIC), où l'interaction avec l'ARN génomique nouvellement produit encapsidé N entraîne un bourgeonnement dans la lumière des compartiments vésiculaires sécrétoires. Enfin, les virions sont sécrétés par la cellule infectée par exocytose. Les étapes clés inhibées par des composés en cours de validation et qui représentent des cibles antivirales attractives sont surlignées en rouge.

Une séquence polyA 3' ; chapeau, structure de chapeau 5' ; ARN db, ARN double brin ; L, séquence de tête ; RdRp, ARN polymérase ARN-dépendante (**Philip, 2020**).

I.4.2.2. Réplication

Les coronavirus expriment et répliquent leur ARN génomique pour produire des copies complètes qui sont incorporées dans des particules virales nouvellement produites. Les coronavirus possèdent des génomes d'ARN remarquablement grands flanqués de régions non traduites 5' et 3' qui contiennent des structures d'ARN secondaire à action cis essentielles à la synthèse de l'ARN. À l'extrémité 5', l'ARN génomique comporte deux grands cadres de lecture ouverts (ORF ; ORF1a et ORF1b) qui occupent les deux tiers du génome coiffé et polyadénylé. ORF1a et ORF1b codent pour 15 à 16 protéines non structurales (nsp), dont 15 composent le complexe de réplication et de transcription virale (RTC) qui comprend, entre autres, les enzymes de traitement et de modification de l'ARN et une fonction de relecture de l'ARN nécessaire au maintien l'intégrité du génome du coronavirus > 30 kb³. Les ORF qui codent pour des protéines structurales et les ORF intercalés qui codent pour des protéines accessoires sont transcrits à partir du tiers 3' du génome pour former un ensemble imbriqué d'ARNm sous-génomiques (ARNm sg). Les protéines accessoires du coronavirus sont des ensembles très variables de protéines spécifiques du virus qui affichent une conservation limitée même au sein des espèces individuelles, mais on pense principalement qu'elles contribuent à moduler les

réponses de l'hôte à l'infection et sont des déterminants de la pathogénicité virale 4,5. Néanmoins, les fonctions moléculaires de nombreuses protéines accessoires restent largement méconnues en raison du manque d'homologies avec les protéines accessoires d'autres coronavirus ou avec d'autres protéines connues 6 (Philip, 2020).

I.4.3. Transmission virale de coronavirus humain

Sur la base des preuves actuellement disponibles, 229E, OC43, NL63 et HKU1 sont bien adaptés aux humains et les virus circulent largement dans la population humaine (Figure 9), la plupart des cas provoquant une maladie bénigne chez les adultes immunocompétents, et aucun de ces virus ne s'est avéré être maintenu dans un réservoir animal. Cependant, dans le cas du SRAS-CoV et du MERS-CoV, les virus ne sont pas aussi bien adaptés pour être maintenus chez l'homme et sont donc probablement disséminés principalement dans le ou les réservoirs zoonotiques, avec des retombées occasionnelles dans la population humaine sensible, peut-être via une espèce hôte intermédiaire (Shuo et al., 2016).

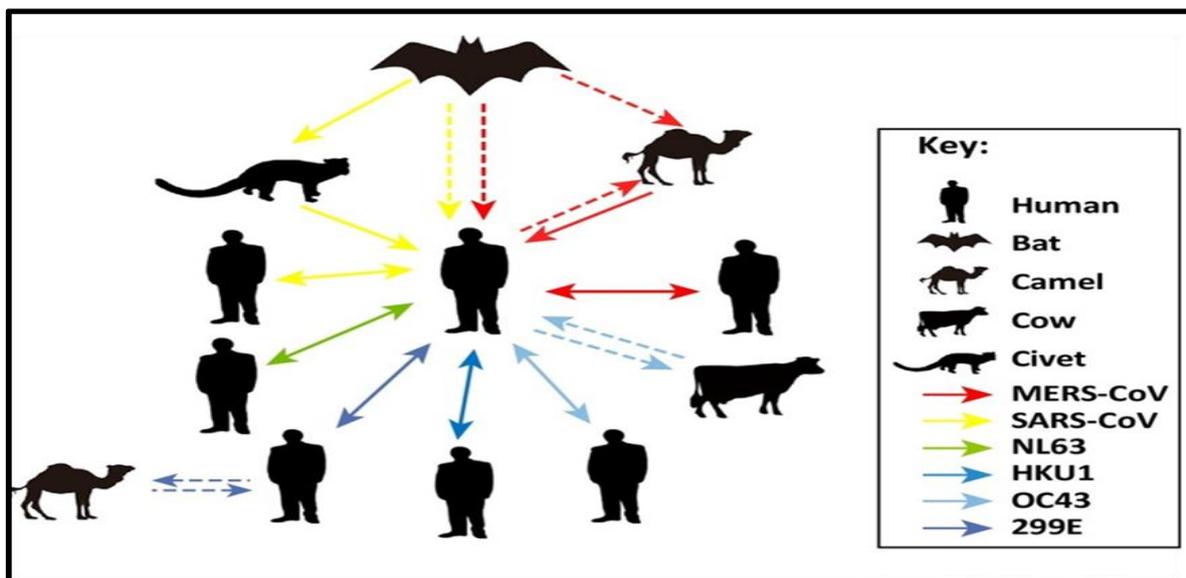


Figure 9 : Transmission intra et inter-espèces des coronavirus humains (Su et al., 2016).

Les flèches rouges, jaunes, vertes, bleues, brunes et violettes représentent la transmission du MERS-CoV, du SARS-CoV, du NL63, du HKU1, de l'OC43 et du 229E, respectivement, entre les chauves-souris, les chameaux, les vaches, les humains et les civettes de palmiers masquées (illustré dans une légende sur le côté de la figure). Les flèches continues représentent une transmission confirmée entre les deux espèces en question, et les flèches brisées représentent une transmission présumée (Su et al., 2016).

I.4.3.1. Transmission virale de SRAS-COV

Les données épidémiologiques impliquaient des civettes de palmier masquées (*Paguma larvata*) provenant des marchés d'animaux vivants (LAM) dans la province du Guangdong, en Chine, comme voie d'exposition au *SRAS-CoV* (Guan, 2003).

Cependant, les civettes de palmier masquées provenant de la nature ou de fermes sans exposition à la LAM étaient largement négatives pour le *SRAS-CoV* (Poon, 2009). Cela suggère que les civettes de palmiers étaient un hôte intermédiaire, mais pas un réservoir pour le *SRAS-CoV* (Perlman et al., 2009). Des études ultérieures ont montré que les chauves-souris sauvages (famille des *Rhinolophidés*), qui peuvent également être trouvées dans la LAM en Chine et servies dans certains restaurants chinois du Guangdong, en Chine, ont des niveaux détectables d'anticorps contre le *SRAS-CoV* et un virus semblable au *SRAS-CoV*. (*SARSr-Rh-Bat CoV*) (Emerg, 2004; kan, 2005) suggérant une origine chauve-souris pour le *SRAS-CoV*. Une relation évolutive entre les coronavirus et les chauves-souris a été proposée, dans laquelle l'ancêtre du *SRAS-CoV* s'est d'abord propagé aux chauves-souris de la famille des *Hipposideridae*, puis aux *Rhinolophidae*, puis aux civettes de palmiers masquées et finalement aux humains (Viro, 2010).

Récemment, deux nouveaux CoV de type SRAS ont été isolés chez des chauves-souris en fer à cheval, et ces virus ont montré la relation la plus élevée avec le *SRAS-CoV* de tous les coronavirus connus de chauve-souris (Ge, 2013), l'analyse ORF8 des CoV de type SRAS chez les chauves-souris suggère que les chauves-souris chinoises sont les réservoirs naturels du *SRAS-CoV* (Dis, 2016), et que des hôtes intermédiaires pourraient ne pas être nécessaires pour une infection humaine directe, en particulier pour certains virus de type *SRAS-CoV* de chauve-souris (Infect, 2016), nous résumons ça dans la figure suivante :

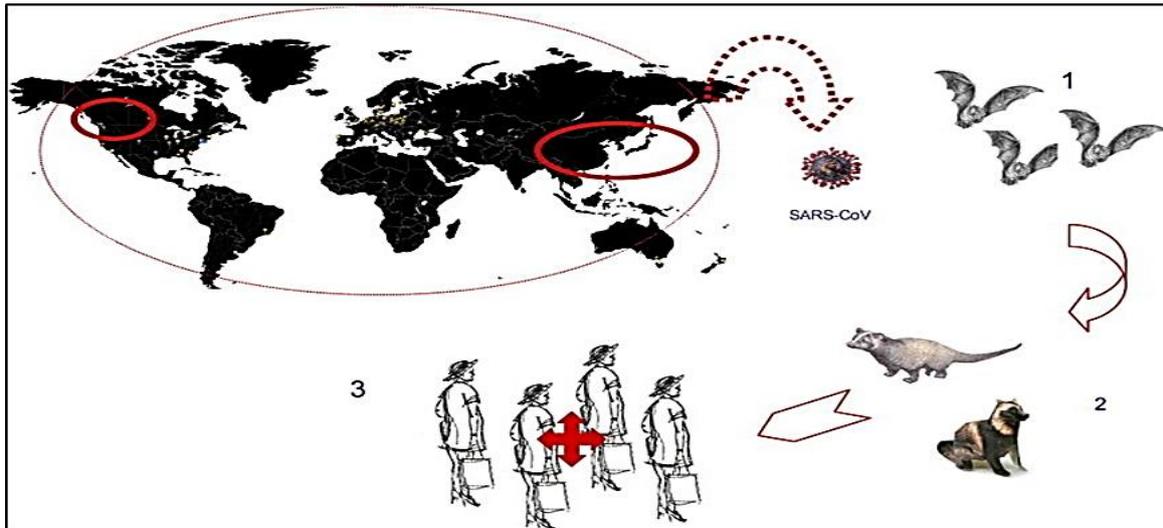


Figure 10 : Épidémie de syndrome respiratoire aigu sévère ou SRAS (Vabret et al., 2009).

I.4.3.2. Transmission virale de MERS-COV

Des virus de type MERS ont été identifiés chez des chauves-souris et pourraient être le progéniteur du virus maintenant reconnu comme endémique chez les chameaux dans certaines parties du Moyen-Orient et d'Afrique. L'infection par le MERS-Cove des chameaux semble provoquer une maladie respiratoire sans mortalité significative chez les chameaux jeunes ou adultes. L'infection par le *MERS-CoV* chez l'homme (Figure11) dont la contamination est proposée par contact avec des chameaux ou du lait ou de la viande de chamelle crue, peut provoquer une pneumonie (point rouge) ou une maladie bénigne ou asymptomatique (point bleu). Actuellement, il semble y avoir une transmission interhumaine limitée du MERS-CoV. Peut provoquer une pneumonie (point rouge) ou une maladie bénigne ou asymptomatique (point bleu). Actuellement, il semble y avoir une transmission interhumaine limitée du MERS-CoV. Peut provoquer une pneumonie (point rouge) ou une maladie bénigne ou asymptomatique (point bleu). Actuellement, il semble y avoir une transmission interhumaine limitée du MERS-CoV (Deng et al., 2021).

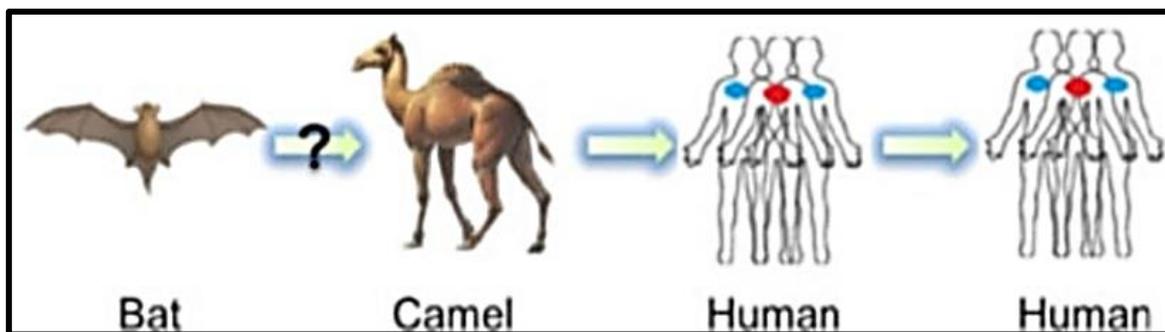


Figure 11: Transmission de coronavirus de type MERS-COV (Deng et al., 2021).

I.4.3.3. Transmission virale de SARS-COV-2

Lorsqu'un virus se propage par transmission respiratoire, il le fait soit avec des virions en suspension sur de grosses gouttelettes, soit avec de fins aérosols expulsés des voies respiratoires du patient primaire. Les gouttelettes sont classiquement considérées comme des particules de plus de 5 µm qui tombent au sol à moins de 6 pieds environ et les aérosols comme des particules de moins de 5 µm qui peuvent rester en suspension dans l'air pendant des périodes prolongées ; cependant, cette dichotomisation peut être une simplification excessive et il est difficile de distinguer la transmission par gouttelettes et par aérosol en milieu clinique (Wei et Li, 2016).

La voie de transmission dominante du *SRAS-CoV-2* est respiratoire (Lu et al., 2020). De plus en plus de preuves indiquent que le virus infectieux peut être trouvé dans les aérosols et dans les échantillons d'haleine expirée (Ma et al., 2020), et il est probable que dans certaines circonstances, y compris pendant les procédures générant des aérosols, en chantant ou dans des environnements intérieurs avec une mauvaise ventilation, le virus peut être transmis à distance par aérosols.

I.5. Identification phylogénétique

I.5.1 Définition de la phylogénie

Le terme « phylogénie » fut inventé par Ernst Haeckel en 1866 pour définir l'enchaînement des espèces animales et végétales au cours du temps. Jusqu'alors le concept était exprimé par le terme « généalogie ». Ce n'est que dans la dernière édition de l'origine des espèces (1872) que Charles Darwin introduisit le mot phylogénie avec la définition suivante : les lignes généalogiques de tous les êtres organisés. Nous définirons la phylogénie comme « le cours historique de la descendance des êtres organisés ».

Haeckel lui-même avait défini la phylogénie comme l'histoire du développement paléontologique des organismes par analogie avec l'ontogénie ou histoire du développement individuel. Les termes « développement » et « évolution » sont tous deux issus de l'embryologie. Pour qualifier les transformations organiques situées dans le temps géologique, le mot « évolution » supplanta progressivement à la fin du XIX^e siècle celui de « développement ». Haeckel fut l'un des artisans de ce succès qui se fit au détriment de « transformisme », terme synonyme « évolutionnisme » et qui reste le plus souvent associé à l'œuvre de J.-B. Monet de Lamarck, quoique ce dernier ne l'utilisât jamais.

Lamarck, en même temps qu'il conçut les bases de la théorie de l'évolution, publia dans sa Philosophie zoologique (1809) un schéma de filiation des animaux. Ce schéma est présenté comme la « distribution générale » des animaux. Ce concept lamarckien exceptionnellement fécond s'oppose explicitement à la classification et se veut une construction qui doit exprimer « l'ordre représentant le plus possible celui même de la nature, c'est-à-dire l'ordre qu'elle a suivi dans la production des animaux et qu'elle a éminemment caractérisé par les rapports qu'elle a mis entre les uns et les autres ». Les « rapports » qu'évoque Lamarck sont les « parentés entre les corps vivants » (**Daru et al., 2008**).

I.5.2. Phylogénie moléculaire

La phylogénie moléculaire consiste à reconstruire les relations de parenté entre les séquences de nucléotides ou d'acides aminés. Elle permet d'étudier l'évolution des séquences et la relation de parenté entre espèces. La nature quantitative des relations, ainsi établies, a permis le développement de méthodes et de règles rigoureuses pour la construction des arbres phylogénétiques (**Cheikh, 2006**). Elle étudie l'histoire évolutive des espèces étudiées à la base d'une portion de leur séquence moléculaire (**Diallo, 2009**).

La phylogénie moléculaire est basée sur les principes suivants :

- Les mutations se produisent au hasard,
- Les mutations s'accumulent au cours du temps,
- Les mutations se produisant chez un individu et se fixent dans sa descendance.

Pour reconstituer des phylogénies, on peut utiliser différents gènes, ou les protéines correspondantes, appelés marqueurs phylogénétiques (**Alauzet, 2009**).

Un marqueur génétique est basé sur la variation dans le génome, cette variation est aussi nommée polymorphisme. Selon le niveau taxonomique étudié, il est essentiel de choisir un type de marqueur qui lui est adapté. Par exemple, si le but visé est de distinguer des groupes d'individus au niveau de l'espèce, les marqueurs utilisés devront être plus conservés que dans le cas des études de populations (Beaulieu, 2007). Les polymorphismes ADN peuvent être détectés dans l'ADN nucléaire et l'ADN d'organites, trouvé dans les mitochondries et les chloroplastes. Les marqueurs moléculaires concernent la molécule d'ADN elle-même, et, comme tels, sont considérés comme des mesures objectives de la variation (**Vicente et al., 2003**).

- 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

- **Marqueurs d'ADN mitochondriaux**

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 de détecter l'hybridation entre les espèces et les sous-espèces d'animaux d'élevage (Nijman *et al.*, 2003) ou même sauvage.

I.5.3. L'arbre phylogénétique

L'arbre phylogénétique est une construction-clé dans l'histoire de la biologie et de la géologie. Son succès opérationnel ne s'est jamais démenti. Imaginons que nous retracions l'histoire de trois espèces choisies arbitrairement (actuelles ou fossiles). En remontant le temps, nous pouvons espérer relier les deux espèces qui dérivent de la même espèce ancestrale ; en remontant plus loin encore, nous rencontrons l'espèce ancestrale des trois espèces. Le dessin en deux dimensions de ces relations de parenté est un arbre composé d'une succession de branchements. Au reste, on n'a jamais trouvé meilleure façon d'illustrer, en deux dimensions, les relations de parenté entre espèces ou groupes d'espèces, en fonction du temps (verticalement) et de la diversité taxinomique (horizontalement). Dans le cas de la diversité, de façon conventionnelle et le plus souvent symbolique, la dimension horizontale permet aussi une représentation du degré de divergence morphologique de deux branches à partir du point de branchement, ce que Darwin nomme la « somme des modifications » (Darwin *et al.*, 2008).

I.6. La phylogénie et l'informatique

I.6.1. Les bases de données

Depuis que les biologistes travaillent avec des séquences en grande quantité, c'est-à-dire depuis le développement et la généralisation de l'utilisation des méthodes rapides de séquençage, la nécessité d'organiser et d'accéder aisément à ces données s'est fait ressentir. Les premières banques de données qui ont été créées concernent les informations structurales sur les protéines puis, très rapidement après, les séquences protéiques et nucléotidiques (**Perrière, 2000**).

Aujourd'hui on peut distinguer les banques dites généralistes des banques spécialistes (**Vergnaud, 2002 ; Pillet, 2000**). Les premières ont pour vocation d'être les plus exhaustives possibles, c'est-à-dire rassembler la totalité des séquences ou informations connues pour tout l'ensemble des espèces avec ou sans expertise particulière. Les secondes se sont constituées autour d'une thématique biologique, afin de réunir les séquences d'une famille protéique pour toutes les espèces. Elles dérivent le plus souvent des banques généralistes et nécessitent l'intervention d'experts qui sont la plupart du temps les auteurs de la banque (**Dorkeld, 1994 ; Blanchet, 1999**).

I.6.2. Les banques généralistes

Les banques généralistes sur les séquences nucléotidiques ou protéiques à accès libre et rapide, sont devenues des outils indispensables pour les chercheurs biologistes. Celles-ci sont régulièrement mises à jour (**Pillet, 2000**).

Dans cette section, nous présentons les principales collections généralistes de séquences nucléotidiques et protéiques ainsi que les centres de saisie qui leur sont associés.

I.6.2.1. Banques de séquences nucléotidiques

D'après (**Perrière, 2000 ; Golding et al., 2003**), il existe de par le monde, trois banques généralistes de séquences nucléotidiques accessibles au public : l'EMBL en Europe, Gênant aux Etats Unis et la DDBJ au Japon.

La base de données EMBL a été créée en 1980 à Heidelberg (**Perrière, 2000 ; Guindon, 2003**). Elle est gérée et maintenue depuis 1994 par l'Institut Européen de Bioinformatique (EBI) à Cambridge. C'est la première source européenne de données sur les séquences nucléiques (**Pillet, 2000**).

GenBank a été mise en place en 1979 au LANL « Los Alamos National Laboratory », à Los Alamos. Depuis 1992, elle est maintenue au NCBI « National Center For Biotechnology Information », à Bethesda (**Perrière, 2000**). Le NCBI a été créé par le Congrès Américain en 1988 et est chargé spécifiquement de développer des systèmes automatisés pour la biologie moléculaire et la biotechnologie (**Pillet, 2000**).

Enfin, la DDBJ est une banque Japonaise de données créée en 1984 et maintenue au NIG à Mishima (**Perrière, 2000**).

Les données présentes dans ces banques proviennent désormais en quasi-totalité de soumissions directes effectuées par les auteurs, par l'intermédiaire du réseau Internet (**Golding et al., 2003**). Et la plupart des revues de biologie moléculaire n'acceptent de publier des articles se référant à des séquences que si celles-ci sont dotées d'un numéro d'accès fourni par les banques. Il convient donc, dès l'obtention d'une nouvelle séquence, de soumettre celle-ci à l'un des trois centres de saisie. Les données restantes sont extraites de la littérature scientifique (essentiellement à partir de documents tels que des livres ou des thèses). Il existe également des procédures de soumissions automatiques pour des séquences provenant des brevets (**Perrière, 2000**).

I.6.2.2. Banques de séquences protéiques

Tout comme leurs homologues consacrés aux séquences nucléotidiques, les banques dédiées aux protéines ont une organisation centrée autour de types d'informations : les annotations biologiques et biochimiques d'une part (avec les références bibliographiques et les données taxonomiques associées) et les séquences d'une autre part (**Perrière, 2000**). TrEMBL et GenPept sont des versions protéomiques des deux grandes banques généralistes de séquences nucléotidiques qui sont : EMBL et GenBank. Ainsi, TrEMBL est constituée de séquences protéiques issues de la traduction des séquences codantes de la banque de séquences nucléotidiques de l'EMBL. De son côté, GenPept, distribuée par le NCBI, correspond à la traduction des parties codantes de GenBank (**Blanchet, 1999**).

Ces deux systèmes ne peuvent cependant être considérés comme de véritables banques protéiques, du fait que les annotations qu'ils contiennent sont uniquement celles provenant des séquences nucléotidiques. Les véritables banques de protéines contiennent de nombreuses informations spécifiques (**Perrière, 2000**).

La principale banque de protéines à l'heure actuelle est sans conteste SwissProt, cette banque a été créée en 1986 par Amos Bairoch à Genève, et elle est maintenue et distribuée conjointement par le SIB et l'EBI (**Blanchet, 1999**). Les séquences présentes dans cette banque proviennent de quatre sources : la traduction des gènes annotés dans EMBL, certaines protéines issues d'autres banques protéiques, la consultation des publications scientifiques et enfin, quelques soumissions directes par les auteurs (**Pillet, 2000**).

Enfin, la banque de séquences protéiques PIR est une des plus anciennes banques de données biologiques puisqu'elle est disponible sur internet depuis 1984. Elle fut mise en place par le NBRF comme succession à « Atlas of protein sequence and structure » maintenu depuis 20 ans par Margaret O. DAYHOFF. Depuis 1988, la banque PIR-International fait l'objet d'une collaboration internationale entre le NBRF et le MIPS et le JIPID (**Blanchet, 1999**).

D'autres banques ayant connu un succès bien moindre que SWISSPROT et PIR ont été également développées. On peut ainsi citer OWL, produite à l'université de Leeds en Angleterre (**Pillet, 2000**).

C'est une banque relativement petite en volume qui n'a pas connu de nouvelles versions depuis 1998 et sa maintenance semble désormais abandonnée. Plus récemment, une nouvelle banque a fait son apparition : il s'agit de KIND, une même banque que OWL mais construite à partir de TrEMBL, GenPept, SWISSPROT et PIR. Là encore, la fréquence d'apparition de nouvelles versions semble assez limitée (**Perrière, 2000**).

Les séquences recueillies par les trois centres de saisie : EBI, NCBI et NIG sont échangées quotidiennement, cela de façon à obtenir un ensemble de données aussi complet et cohérent que possible. La conséquence de cela est qu'en pratique, ces trois banques n'en font qu'une car leur contenu est identique à plus de 99,9%. Ceci amène régulièrement les responsables des centres de saisie à se poser la question de l'utilité de la maintenance de trois banques différentes. Il existe ainsi depuis longtemps un projet de fusion d'EMBL et GenBank et de la DDJB en un seul système (**Golding et al., 2003**).

Depuis leur création, la taille de ces banques n'a pas cessé de croître. Si l'on regarde l'allure de cette croissance, on constate qu'elle peut être approximativement assimilée à une exponentielle pour les séquences nucléotidiques. Par contre, le nombre de séquences protéiques croît beaucoup moins vite. D'après (**Perrière, 2000**), l'accélération constatée en 1995 correspond au développement des techniques de PCR, largement utilisée pour l'étude des génomes eucaryotes.

I.6.3. Les banques spécialistes

Du fait de l'augmentation exponentielle du nombre de séquences nucléotidiques et protéiques publiées, la nécessité de développer des banques spécialisées est rapidement apparue. Ainsi, aujourd'hui, il en existe de nombreuses. Les informations disponibles dans ces banques et leur mise à jour sont très variables (**Blanchet, 1999**).

D'après (**Perrière, 2000**), on distingue deux types de banques spécialisées : les banques thématiques et les banques génomiques.

Les banques thématiques se consacrent à une thématique biologique bien précise (par exemple : banques sur les structures moléculaires des protéines, des complexes protéines-acides nucléiques, de virus et de polysides, banques sur les structures de l'ARN, banques sur des familles de gènes...etc.). Ces banques intègrent donc des séquences et des données qui sont spécifiques à ce domaine. On peut par exemple citer :

La base PDB est la plus connue des banques des données sur les structures moléculaires. Elle rassemble les structures tridimensionnelles des macromolécules obtenues par cristallographie aux rayons x. Elle contient aussi des références bibliographiques et des données sur la structure primaire et secondaire des protéines (**Golding et al., 2003**).

La RDP est une banque américaine qui contient non seulement les séquences des ARN ribosomiques de la grande et de la petite sous-unité du ribosome, mais aussi des alignements et des arbres phylogénétiques (**Perrière, 2000 ; Guindons, 2003**).

Les bases de données génomiques sont des banques de données centrées sur les génomes des organismes étudiés en biologie, c'est-à-dire l'ensemble des séquences d'ADN contenues dans les chromosomes d'organismes spécifiques. Ces organismes modèles peuvent être des virus, bactéries, champignons, plantes et animaux. En plus des données de séquences, ces bases rassemblent aussi d'autres types de données (description de phénotypes, mutations, expression de gènes, bibliographies...etc.) (**Pillet, 2000**). Dans cette catégorie on peut citer :

- ✓ ECDC est une banque qui présente une compilation des séquences d'E. Coli contenues dans EMBL et dans Génant.
- ✓ La base SGD « Saccharomyces Génome Database » compile les séquences nucléiques des gènes codant pour les protéines de la levure *Saccharomyces cerevisiae*.

- ✓ PGD « Plant Genome Database » est une base de données qui compile des données sur le génome de plusieurs plantes céréalières. Par ailleurs, ATDB « Arabidopsis thaliana Datamasse » rassemble toutes les informations sur les données génétiques, le clonage et le séquençage des molécules de la plante crucifère Arabidopsis thaliana.

I.7. Les logiciels utilisés dans la phylogénie

I.7.1. Présentation des logiciels d'analyse phylogénétique

Aujourd'hui, il existe une myriade de logiciels, implantés sur différentes machines depuis le microordinateur jusqu'à des ordinateurs les plus puissants, écrits dans tel ou tel langage, traitant tel ou tel problème.

Dans le domaine de la reconstruction phylogénétique, il y a plusieurs programmes traitant les différents aspects de la phylogénie, allant des problèmes de comparaison des séquences, calcul des distances évolutives, reconstruction et visualisation d'arbres phylogénétiques et le Nettoyage de l'alignement. Pour cela nous allons décrire quelques grandes catégories de logiciels utilisés pour les analyses phylogénétiques selon leurs fonctions.

I.7.1.1. Les méthodes de reconstruction phylogénétique et les logiciels utilisés

Les méthodes de reconstruction phylogénétique reposent presque toujours sur l'optimisation d'un critère permettant de comparer les phylogénies possibles d'un ensemble de taxons, nous décrivons les principales méthodes de reconstruction phylogénétique, les méthodes de distances, de parcimonie et du maximum de vraisemblance.

I.7.1.1.1. La méthode de parcimonie

Les méthodes de parcimonie reposent sur l'idée que l'histoire évolutive la plus plausible est celle qui nécessite un minimum de mutations. Ces méthodes recherchent donc parmi tous les arbres possibles, et toutes les séquences possibles aux nœuds internes, la combinaison qui minimise le nombre total de mutations requis pour expliquer les données observées.

Les différentes mutations peuvent avoir des coûts différents, et il est possible d'attribuer des poids différents à chacun des sites.

Dans cette approche, les sites peuvent être traités de manière indépendante. Le nombre de mutations correspondant à un jeu de séquences particulier est égal à la somme des mutations

requis par chacun des sites. Les coûts attribués aux mutations sont généralement symétriques (Renwez et al., 2003).

Ces parcimonies sont réalisées par des logiciels spécialisés tels que : PHYLIP (Phylogeny Inference Package), PAUP, PAML, EDIBLE, PTP, PHYML, PARDOOT, ALIFRITZ, CRUX, PRAP.

I.7.1.1.2. La méthode Bayésien inference

La méthode de bayes, plus exactement appelée méthode Bayesian Markov Chain Monte Carlo (MCMC), a récemment apporté un nouveau moyen de construire l'arbre phylogénétique le plus probable. Cette méthode semble donner des résultats prometteurs et son emploi s'est dernièrement généralisé.

Le principe de cette méthode est relativement simple : une série d'arbre est construite sur le principe que chaque arbre, très voisin du précédent, doit apporter une amélioration à la construction générale. Ainsi, un premier arbre phylogénétique est construit au hasard. Puis un élément de cet arbre est changé, également au hasard. Ce nouvel élément peut être la longueur d'une branche, ou le changement de position d'un taxon puis les arbres successifs sont conservés selon un algorithme particulier (dit de metropolis –Hastings) qui accepte les arbres au *prorata* de leur apport : si le nouvel arbre apporte une amélioration, il est toujours accepté ; s'il est moins bon de 10% de l'arbre précédent, il n'est accepté que 90% du temps (Pagel, 2002).

Cette méthode est réalisée par des logiciels spécialisés tels que : PAML, BAMBE, PAL, Vanilla, MrBayes, Mesquite, PHASE, BEAST, MrBayes tree scanners, P4, SIMMAP, IMA2, BALI-Phy.

I.7.1.1.3. La méthode de Maximum de vraisemblance

Les méthodes de maximum de vraisemblance utilisent un modèle mathématique du processus d'évolution des séquences pour définir la probabilité qu'une phylogénie puisse produire les séquences observées. Elles cherchent ensuite la phylogénie pour laquelle cette probabilité est maximale. Pour utiliser une méthode de maximum de vraisemblance, il faut donc être capable de choisir un modèle d'évolution, d'estimer ses paramètres, et de calculer la vraisemblance d'un arbre pour ce modèle (Ranwez, 2013).

Cette méthode est réalisée par des logiciels spécialisés tels que : PHYLIP, DAMBE, EMBOSS, PRAP, PAUP*, ARB, Astrée, Creux, Vanilla, McRate, DART.

I.7.1.2. Les logiciels de Nettoyage de l'alignement

C'est l'élimination des régions divergentes et présentant peu de signal phylogénétiques (peu informatives). Ces régions peuvent ne pas être homologues ou avoir été saturées par des substitutions ; pour réaliser cette élimination il faut utiliseré Séquencher ou Bio Edit.

Bio Edit est un éditeur d'alignement de séquences et un programme d'analyse de séquences piloté par la souris, facile à utiliser, conçu et écrit par un étudiant diplômé qui sait combien il est frustrant et consommateur de temps de compter sur les processeurs de mots et les programmes de ligne de commande pour la manipulation de séquence. Bio Edit est destiné à fournir un programme unique qui peut gérer les fonctions d'édition et de manipulation de séquences et d'alignement les plus simples que les chercheurs sont susceptibles de faire quotidiennement, ainsi que quelques analyses de séquences de base.

Sequencher est un logiciel de bio-informatique produit par la société Gene Codes Corporation. Ce logiciel permet de faire de l'assemblage de plusieurs séquences d'ADN contigües relativement courtes afin de créer des séquences plus longues (appelées contigs).

Le logiciel accepte les électrophorègrammes des différents contigs issus des séquenceurs de gène et les aligne (appariement), ce qui permet de les comparer les uns aux autres afin de produire une séquence corrigée portant sur l'ensemble de la portée des contigs mis bout à bout.

I.7.1.3. Les logiciels pour l'alignement de séquence

L'alignement est une opération qui permet de comparer des séquences de nucléotides pour repérer les éléments correspondants. Cette opération est nécessaire pour toutes les comparaisons de séquences (**Faurnel et al., 1834**).

Ces alignements sont réalisés par des logiciels spécialisés tels que : MEGA, MUSCLE, TreeAlign, clustal W, DAMBE, T.coffee, BIRCH, EMBOSS.

I.7.1.4. Les logiciels de traitement de séquence

Quelques logiciels généralistes gratuits et téléchargeables pour le traitement des séquences biologique.

Séquence suite manipulation : ensemble de pages html et java script permettant des analyses simples sur les séquences nucléotidiques et protéiques, très bien fait et facile utiliser (mac, pc, linux).

- FastPCR : design des amorces pour PCR ; il faut lire l'aide en ligne pour une prise en main du logiciel ; il permet la prédiction et l'optimisation des amorces.
- MB DNA Analysis : package permettant de réaliser les cartes de restriction, %GC, recherche d'ORF, dot-plot, alignement multiple, calcul de masse moléculaire, prédiction de structure secondaire.
- DRAW32 : logiciel de dessin de vecteur. Il réalise également des cartes de restriction et propose une électrophorèse virtuelle des fragments obtenus.

Antheprot : dédié à l'analyse des séquences protéiques.

I.7.1.5. Logiciels de visualisation et reconstruction d'arbres phylogénétiques

Historiquement, les premières personnes ayant démocratisé la visualisation sous forme d'arbre sont les généalogistes et les naturalistes. De plus, Charles Darwin les réutilisa allègrement en 1859 dans son ouvrage devenu référence *De l'origine des espèces*. Mais si l'on cherche le précurseur historique, c'est en réalité Willi Hennig, père de la cladistique, qui a réussi à démocratiser les arbres phylogénétiques.

Comme en généalogie où les arbres sont utilisés pour visualiser les relations de parenté, les arbres sont un très bon moyen de faire ressortir l'évolution d'une ou plusieurs espèces.

Actuellement, Il existe plusieurs programmes qui permettent la visualisation d'arbres phylogénétique, par exemple, les programmes : PhyloDraw, NJplot, GeneTree, Phylip (DRAWTREE et DRAWGRAM), Genedoc, Dambe, Treecon, TreeView, et Spectrum (**Choi et al., 2000**).

Objectif de notre étude:

Notre objectif est d'inventorier les espèces touchées par le coronavirus afin de faire un analyse phylogénétique comparative entre les différentes séquences d'ADN qui sont en commun entre l'homme et les autres espèces animales. En utilisant plusieurs logiciels bioinformatique.

La réalisation de ce travail est faite 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.

Cette étude se base sur l'utilisation des outils bioinformatique dans les analyses phylogénétiques de la famille de coronaviridae, tel que la reconstruction de l'arbre phylogénétique, identifier des haplotypes, et inventorier les espèces touchées par le coronavirus et faire comparaison entre les souches qui ont touché l'être humain et les autres souches.

Dans ce chapitre nous allons présenter quelques fonctions des logiciels utilisés dans notre travail.

II.1. Analyses moléculaires et phylogénétiques

L'objectif de cette partie repose essentiellement sur l'utilisation de plusieurs logiciels de bioinformatique afin d'étudier une base de données de biologie moléculaire pour tracer l'arbre phylogénétique et faire comparaison entre la souche qui a touché l'être humain et les autres souches.

Les spécimens cliniques utilisés sont des échantillons respiratoires LBA, écouvillons nasopharyngés dans des milieux de transport UTM, pour cette raison nous avons la procédure d'extraction de l'ARN viral qui comprend 4 étapes :

-La lyse des cellules contenues dans le spécimen nasopharyngé ou respiratoire, la liaison des acides nucléiques aux particules magnétique, le lavage et l'élimination des débris cellulaires, l'éluion des acides nucléiques. Cependant l'amplification de l'ARN viral se fait par RT-PCR (Réaction de Polymérisation en Chaîne par Transcription inverse), et elle est mise au point pour utiliser les ARN comme matrice d'amplification de la PCR qui conduit à la synthèse de l'ADNc à fin d'être séquencer par la méthode utiliser aujourd'hui "**Sanger**" consiste sur l'utilisation des didésoxyribonucléotides (**Marc, 2020**).

II.2. Extraction des données

II.2.1. Définition de la GenBank

La GenBank est une base de données construite et distribuée par le National Center for Biotechnology Information (NCBI). Elle est une collection de toutes les séquences de nucléotides et de protéines et d'annotations bibliographiques et biologiques disponibles au public dont son URL (**Sayers et al., 2021**). Ces enregistrements sont générés à partir des

communications directes avec la séquence d'ADN des bases de données des auteurs originaux qui donnent leurs dossiers pour rendre les données à la disposition de tout le monde ou de le faire dans le cadre du processus de publication (Clark et al., 2016).

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

1. Construire l'URL suivante : <https://www.ncbi.nlm.nih.gov/genbank/>
2. Choisir recherche par taxonomie sur la source NCBI.
3. Ecrire le nom de la famille « *Coronaviridae* » (Figure 9).

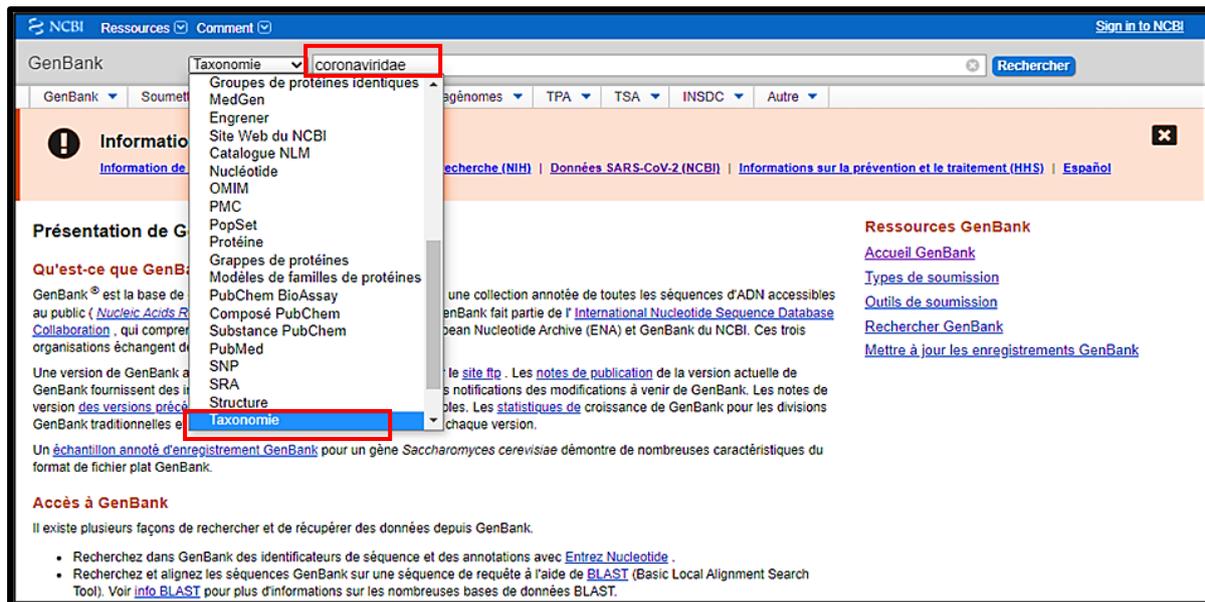


Figure 12 : Les étapes initiales pour récupérer les séquences de la GenBank

4. En entrant sur *Coronaviridae*, nous obtenons toutes les souches virales qui appartiennent à cette famille (Figure 13 a, b).

The screenshot shows the NCBI Taxonomy Browser for 'Coronaviridae'. The search bar contains 'Coronaviridae'. The taxonomic information is displayed, including the Taxonomy ID (11118), current name (Coronaviridae, ICTV accepted), and NCBI BLAST name (viruses). A table of 'Entrez records' is shown on the right, listing various databases and their associated counts.

Database name	Subtree links	Direct links
Nucleotide	1,273,329	1
Protein	6,263,826	2
Structure	1,796	-
Genome	53	-
Popset	1,086	7
Conserved Domains	256	123
GEO Datasets	833	-
PubMed Central	19,370	8,465
Gene	969	-
SRA Experiments	974,311	23
Identical Protein Groups	663,969	2
Bio Project	293	6
Bio Sample	1,506,533	6
Assembly	231	-
Genetic Testing Registry	80	-
Probe	300	-
PubChem BioAssay	292	4
Taxonomy	1,756	1

Figure 13 a : Tableau informatif de *Coronaviridae* obtenue par la GenBank.

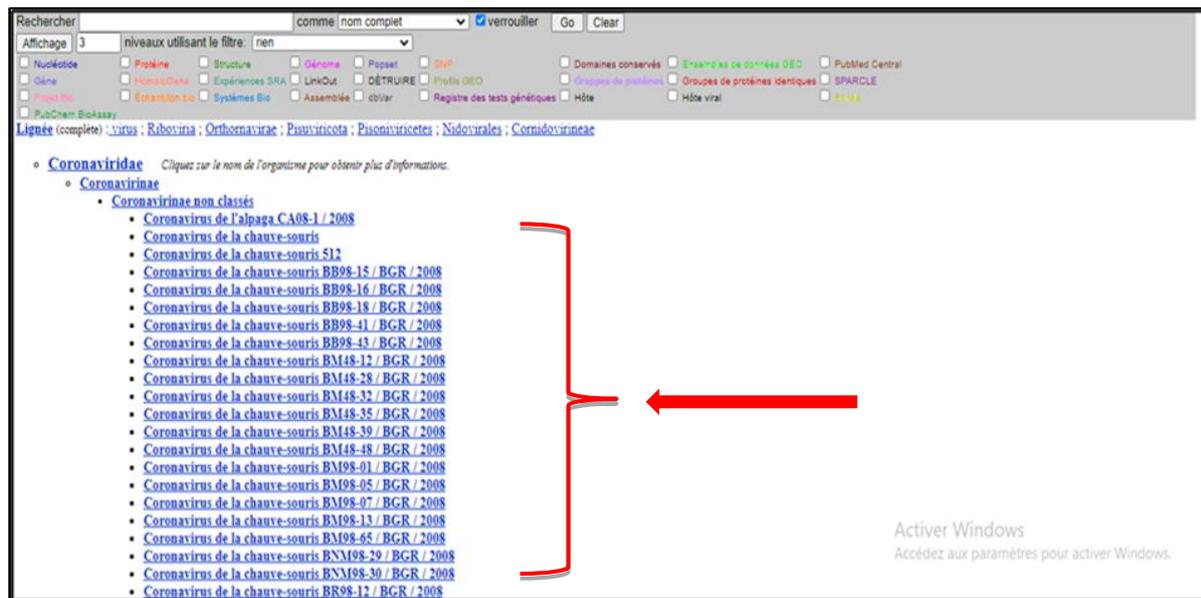


Figure 13 b : Quelques souches virales de *Coronaviridae* non classée obtenues dans la GenBank qui sont indiquées par la flèche rouge.

5. Nous avons analysé toutes les souches une par une pour identifier le nombre des séquences nucléotidiques trouvées dans cette banque (**Figure 14**).

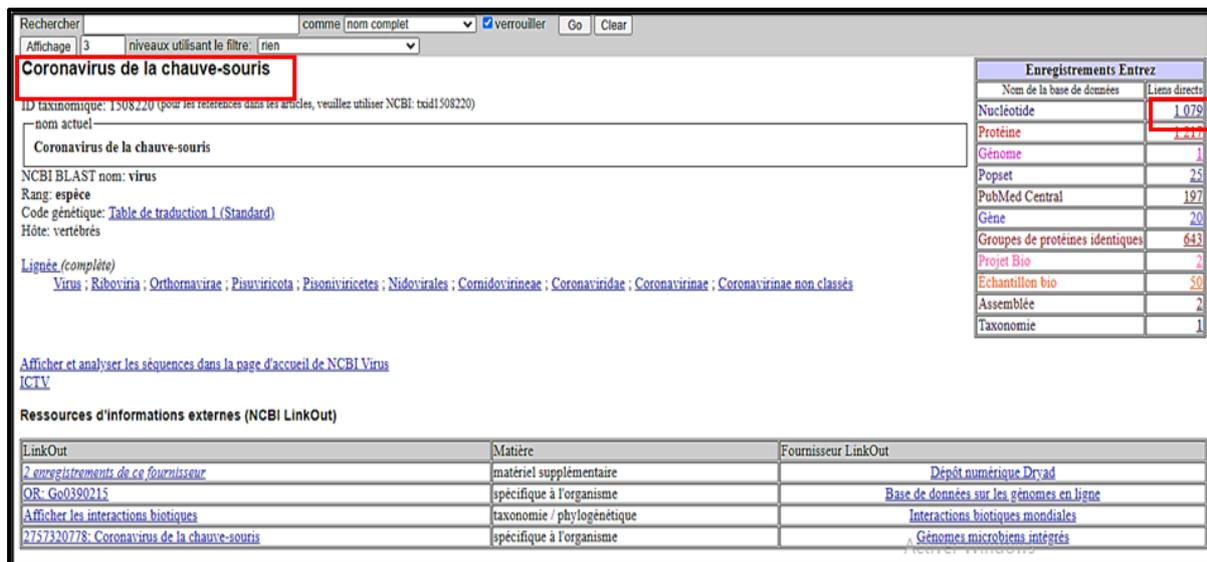


Figure 14 : Exemple des séquences nucléotidiques trouvées pour coronavirus non classé de la chauve-souris.

Au total, la GenBank contient 1082 séquences de l'espèce de chauves-souris de la famille de coronavirusidae.

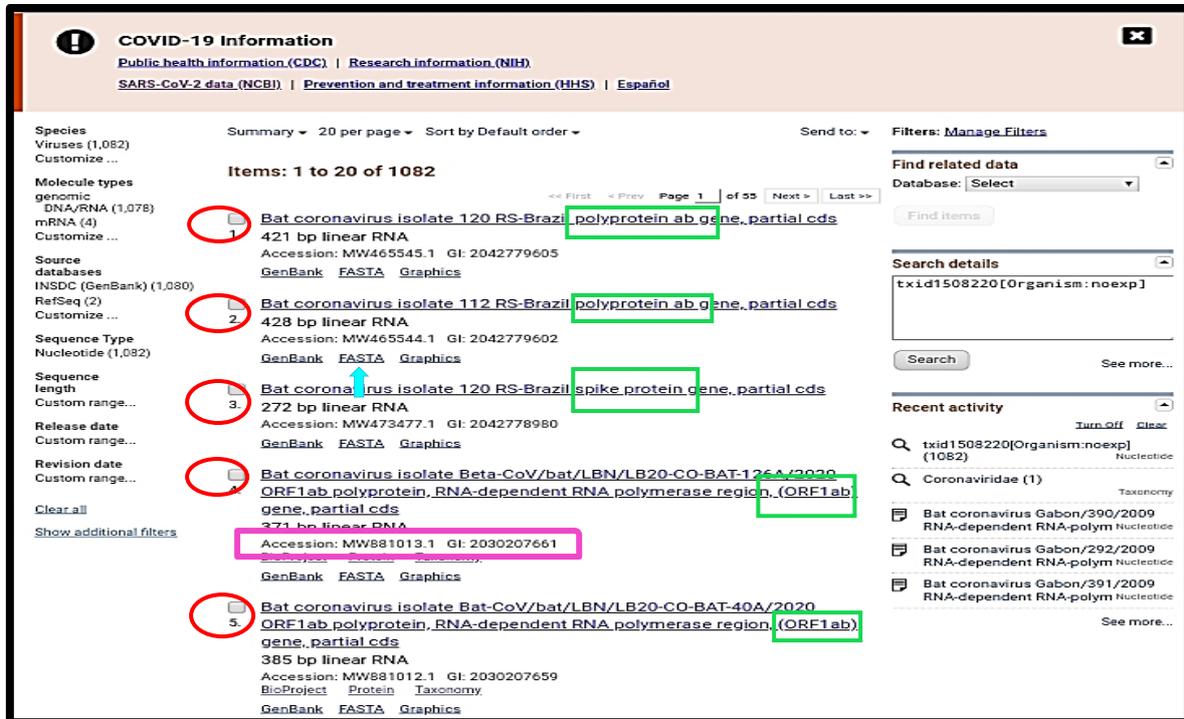


Figure 15 : Exemple de séquence nucléotidique de coronavirus de chauve-souris.

-Les séquence nucléotidique sont indiqués par des cercles rouges.

-Les gènes sont indiqués par des carrés verts.

-Pour accéder au format FASTA il faut cliquer sur l'endroit indiqué par la flèche bleue.

-le code GI et (le locus .1) ont le même rôle: est l'accèsion au séquences. Représentées par le carré en rose.

6. Les séquences représentatives de chacune de toutes les souches ont été téléchargées sous format FASTA à partir de la "GenBank" (**Figure 16**).

```

FASTA - Envoyer à -
Isolat de coronavirus de chauve-souris Beta-CoV / bat / LBN / LB20-CO-BAT-126A
/ 2020 ORF1ab polyprotéine, région ARN-polymérase ARN-dépendante, gène
(ORF1ab), CD partiel
GenBank: MW881013.1
Graphiques GenBank
> MW881013.1 Isolat de coronavirus de chauve-souris Beta-CoV / chauve-souris / LBN / LB20-
CO-BAT-126A / 2020 ORF1ab polyprotéine, région ARN polymérase ARN-dépendante, gène
(ORF1ab), CDS partiels
GCATCACTTGTGCTTGTGCTCGCAACATAGTACTTGTGTAAACCTTTCACACCGTTTCTACAGGTTAGCTA
ATGAGTGTGCTCAGGTAAGTAAATGGTTATGTGTGGCGGTTCACTTTATGTGAAACCAAGGGGTAC
ATCTTCAGGAGATGCCCACTGCTTATGCTAATAGTGTCTTAAACATTTGTCAAGCTTTACAGCTAAT
GTTAATGCACCTTTGTCTACTGATGGTAACAAAATTGCCGACAAGTATGTCGCAATTTACAACATAGAC
TTTATGAATGTCTCTACAGSAATAGAGACGTTGATCATGAATTTGTAGAGGAATTTACGCTTATCTGCG
TAAACACTTTTCTATGATGAT

```

Figure 16 : Exemple d'une séquence nucléotidique au format FASTA.

La base de données a été construite pour sélectionner les isolats à étudier et générer les fichiers de sortie correspondant au format informatique « Fasta ». Ces fichiers peuvent alors être directement utilisés dans nombre logiciels de traitement et d'analyse de séquences.

II.2.1.1. Traitement des séquences de la GenBank :

Pour le traitement des séquences, il existe plusieurs logiciels permettant de générer des arbres phylogéniques environ 367 logiciels de bioinformatique, parmi eux : PHYLIP, PAUP*, MEGA, Phylo_win, ARB, DAMBE, PAL, Bionumerics, Mesquite, PaupUp, BIRCH, Bosque, EMBOSS, phangon bio, ETE, DendroPy, SeaView, Crux.

- Nous avons choisi FaBox, DnaSP6 MEGA5, DAMBE, Network10.2

II.2.1.1.1. MEGA :

Moléculaire Génétique évolutive Analyse (MEGA) est un programme qui permet explorer et d'analyser des séquences nucléidiques et protéines. En effet, dans l'études d'évolution moléculaire il est souvent nécessaire de savoir certaines propriétés statistiques de base telles que les fréquences des différents nucléotides, les fréquences des codons et les rapports de transitions transversions. Il permet également le calcul des distances évolutives entre les séquences sous différents modèles de substitution (Jukes et Cantor, Kimura à 2 paramètres et Tamura-Nei) et la reconstruction des arbres phylogénétiques selon trois méthodes, UPGMA Neighbor-Joining et Parcimonie (Kumar et al., 2001).

Nous pouvons alors procéder à l'analyse phylogénétique à partir des séquences nucléotidiques des souches virales de la famille « *Coronaviridae* » par différentes étapes et programmes.

II.2.1.1.2. FABOX :

Le programme FaBox a été employé pour convertir les fichiers, de la page principale (**Figure 17**) l'utilisateur peut choisir une liste de services, chacune effectuer une tâche simple. En outre, il y a des liens aux services de conversion où l'utilisateur peut convertir n'importe quel format de données en format de Fasta.

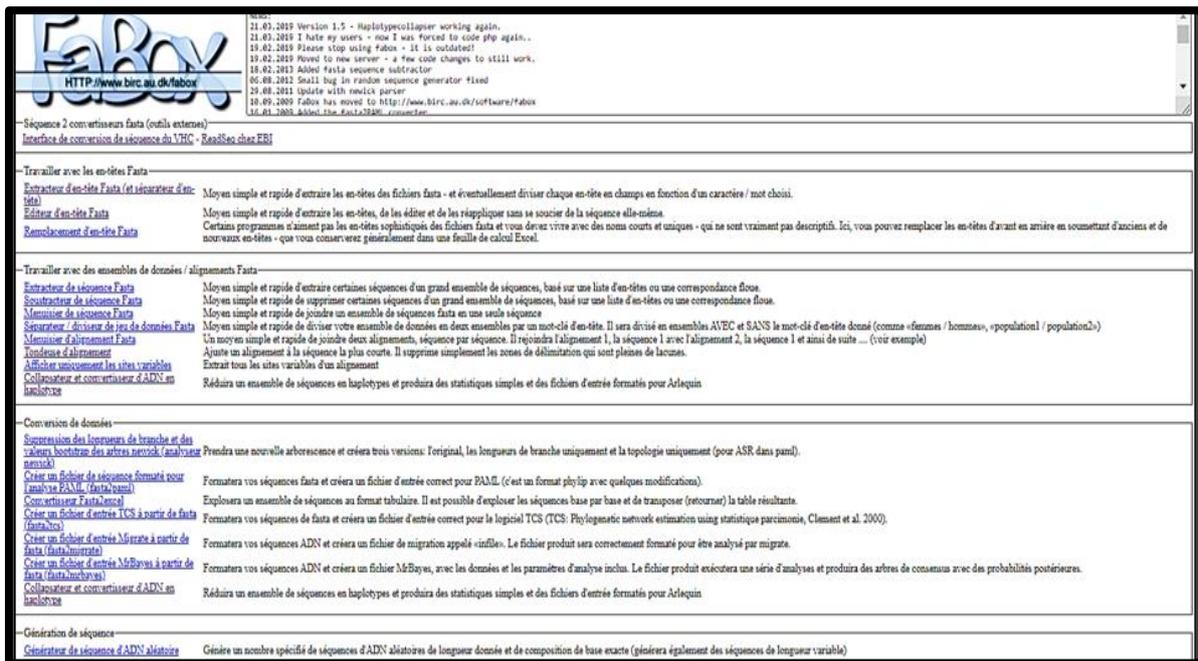


Figure 17 : La page principale du programme Fabox.

L'utilisateur peut entrer des données en téléchargeant un fichier de données ou simplement en collant des données directement dans le web browser.

On s'attend à ce que la boîte à outils se développe sur la base des demandes des service particuliers et des convertisseurs à l'avenir (**Figure 18**).

Figure 18 : La conversion des séquences sous forme Fasta par l'utilisation de Fabox.

Alignement cropper va réduire l'alignement (en format Fasta) à la longueur de la séquence la plus courte (suppression des gaps dans le début et la fin de l'alignement). Appuyez sur " show example" pour l'essayer sans soumettre votre propre séquence (**figure 19**).

Figure 19: Alignment trimmer (la forme Fasta).

Après l'intégration des séquences représentatives des souches virales de coronavirus dans le programme MEGA5 sous format FASTA, un alignement multiple a été réalisé au moyen du

logiciel MEGA version 5 (Tamura et al., 2011) (Figure 20), en utilisant le programme ClustalW.

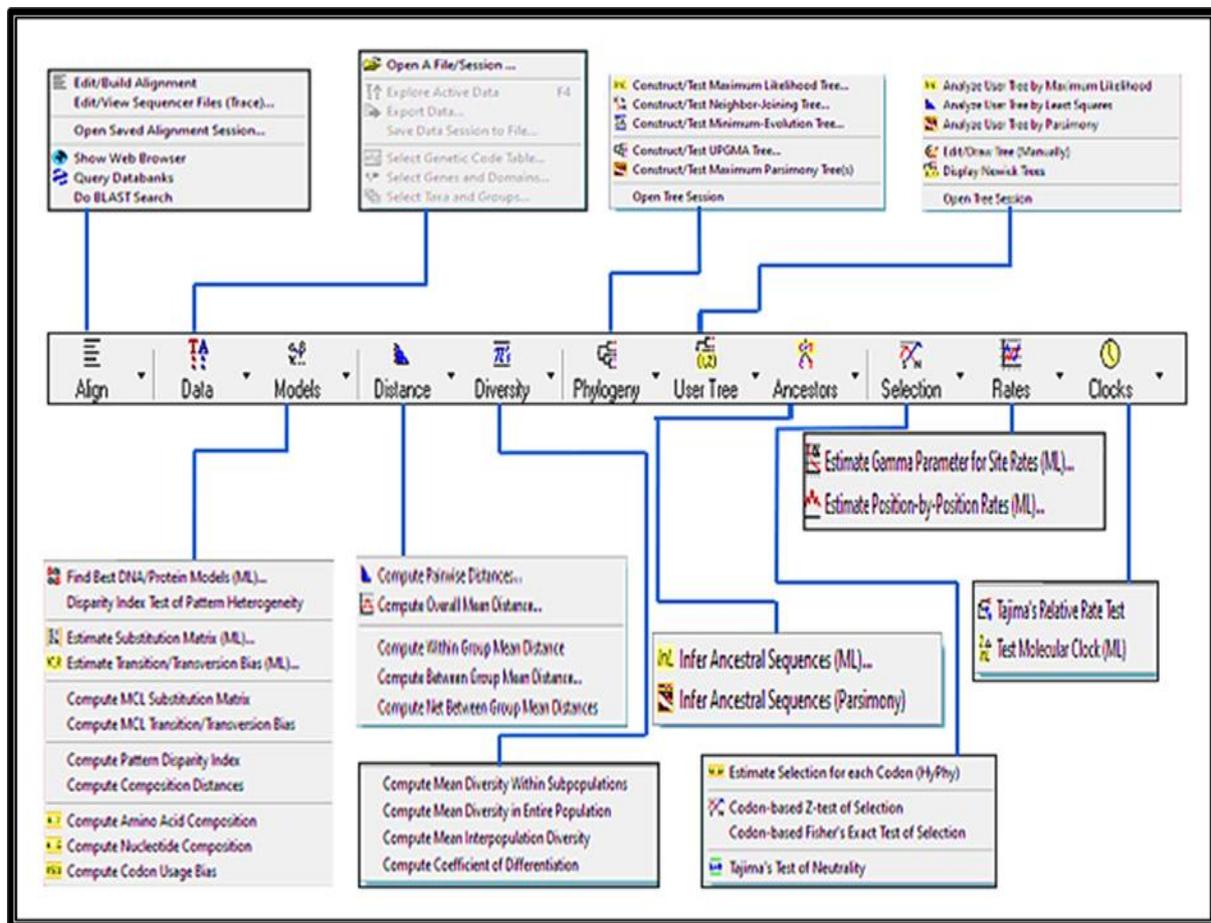


Figure 20 : Barre d'outils de MEGA 5.

Nous choisissons ClustalW (figure 21) c'est un programme d'alignements multiples des séquences d'acides nucléiques ou protéiques. Ces séquences sont fournies par l'utilisateur ou proviennent des banques de données. Ce type d'alignement est très utile puisqu'il permet de comparer des séquences provenant de différentes sources et d'étudier les différences et les similitudes entre les séquences de même type mais appartenant à des groupes différents. Ceci permet d'identifier les régions qui sont conservées d'une espèce à une autre (Thompson et al., 1994).

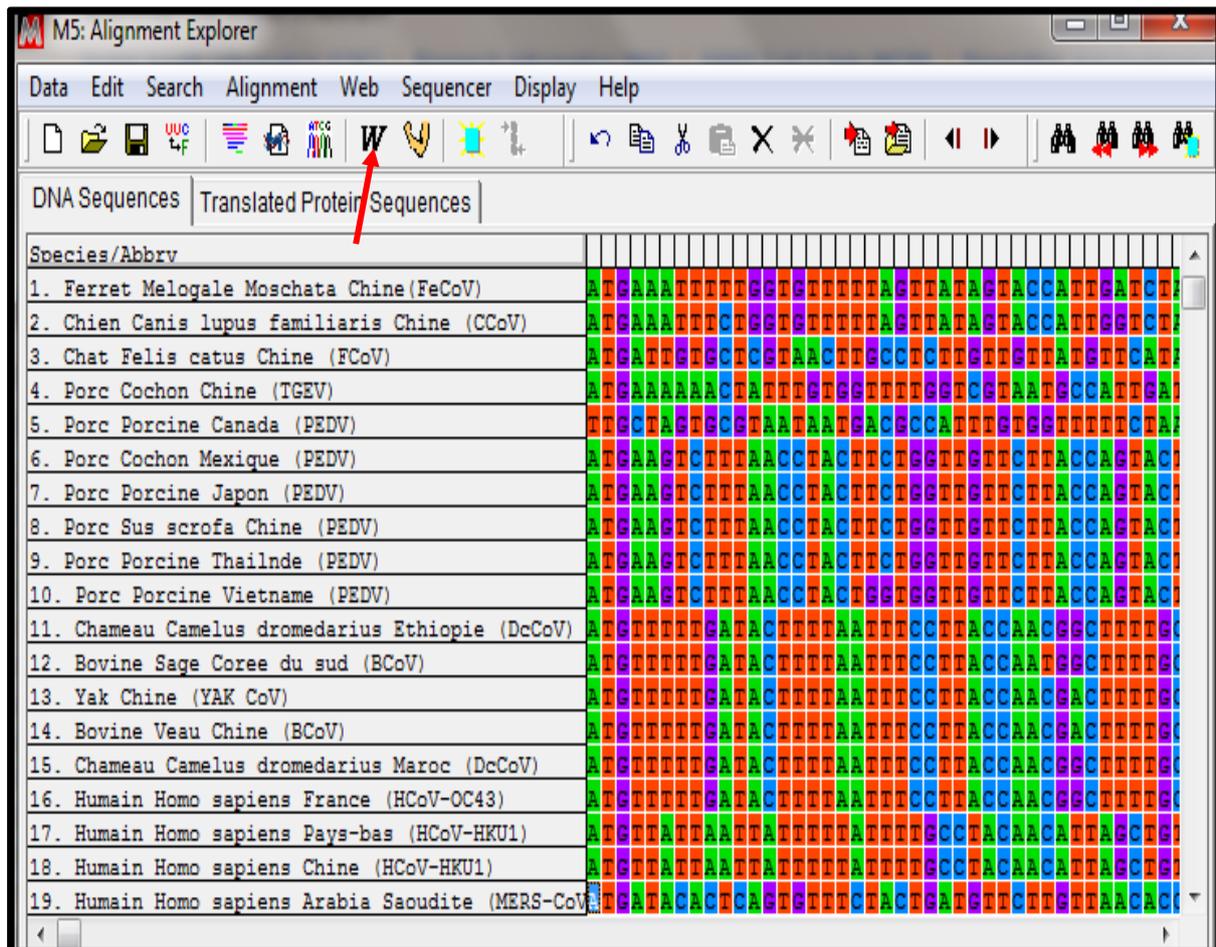


Figure 21 : Le programme ClustalW utilisé pour l'alignement multiple des séquences du marqueur moléculaire choisi

- Une fois les séquences alignées elles sont ensuite comparées et servent à la construction d'arbres phylogénétiques. Plusieurs méthodes ont été développées pour construire un arbre phylogénétique (**Golding et Morton, 2003**), ces méthodes se basent sur la mesure des distances entre les séquences (c'est à dire le nombre de substitutions de nucléotides ou d'acides aminés entre chaque paire de séquences) (Cheikhrouhou, 2006) parmi ces méthodes la méthode de Neighbor joining (NJ), Maximum Likelihood, Maximum Parcimony et UPGMA,
- La méthode UPGMA (Unweighted Pair Group Method of Arithmetic Averages) qu'on a utilisé dans le logiciel MEGA 5(**figure 22**), afin de construire notre arbre phylogénétique. C'est une méthode très simple et très rapide, développée originellement pour construire des phénogrammes taxonomiques (**Golding et Morton, 2003**).

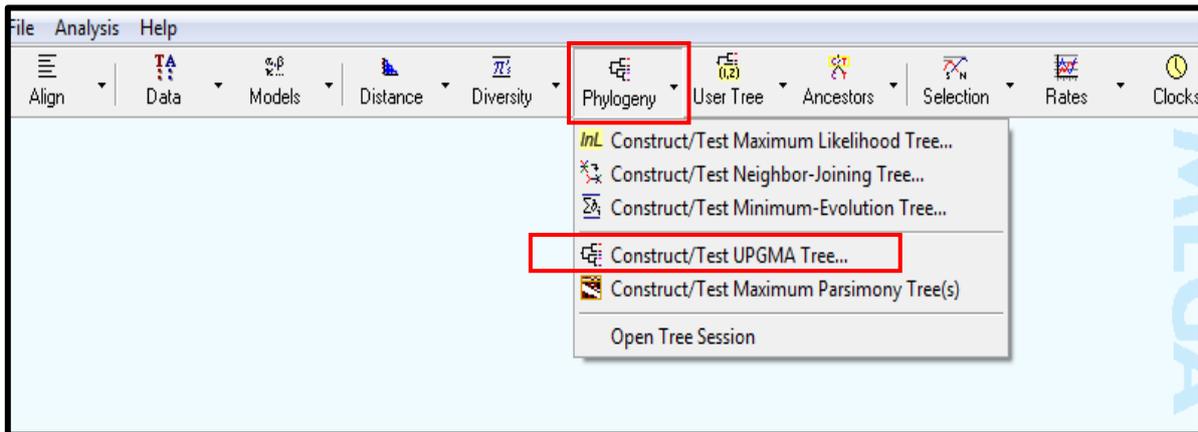


Figure 22 : La méthode UPGMA utilisée pour la construction de l'arbre phylogénétique

La fiabilité des branches des arbres construits a été évaluée à l'aide de la méthode "Bootstrap » (figure 23). Les valeurs de "Bootstrap" pour 1000 répliques sont ramenées en pourcentages et ils sont indiqués au niveau des nœuds.

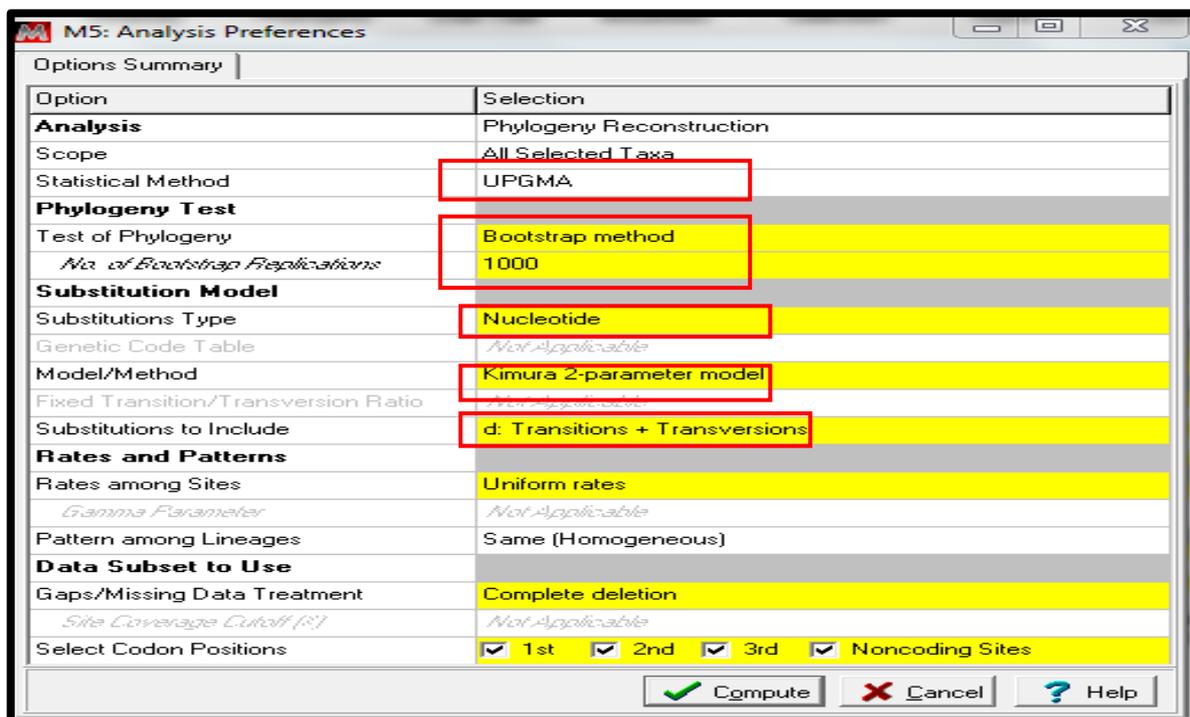


Figure 23 : Les paramètres utilisés pour la reconstruction phylogénétique.

II.2.1.1.3. DNASP :

Est un logiciel pour l'analyse du polymorphisme de nucléotide des données alignées d'ordre d'ADN. Il peut estimer plusieurs mesures de variation d'ordre d'ADN dans et entre les populations (dans des sites non codants, ou dans diverses sortes de positions de codons) ainsi que le déséquilibre de liaison, la recombinaison, le flux génique et les paramètres de conversion génique. DnaSP peut également effectuer plusieurs tests de neutralité. En plus il peut estimer les intervalles de confiance de certaines statistiques d'essai par le coalescent. Les résultats des analyses sont présentés sous forme de tableau et de graphique. La version 5 met en œuvre un certain nombre de nouvelles fonctionnalités et des méthodes analytiques permettant de nombreuses analyses du polymorphisme de l'ADN sur de grands volumes de données.

DnaSP (**Figure24**) accepte cinq formats de fichier de données d'entrée : FASTA, MEGA, NBRF / PIR, NEXUS et PHYLIP. Les résultats des analyses sont présentés sous forme de tableau et de graphes.

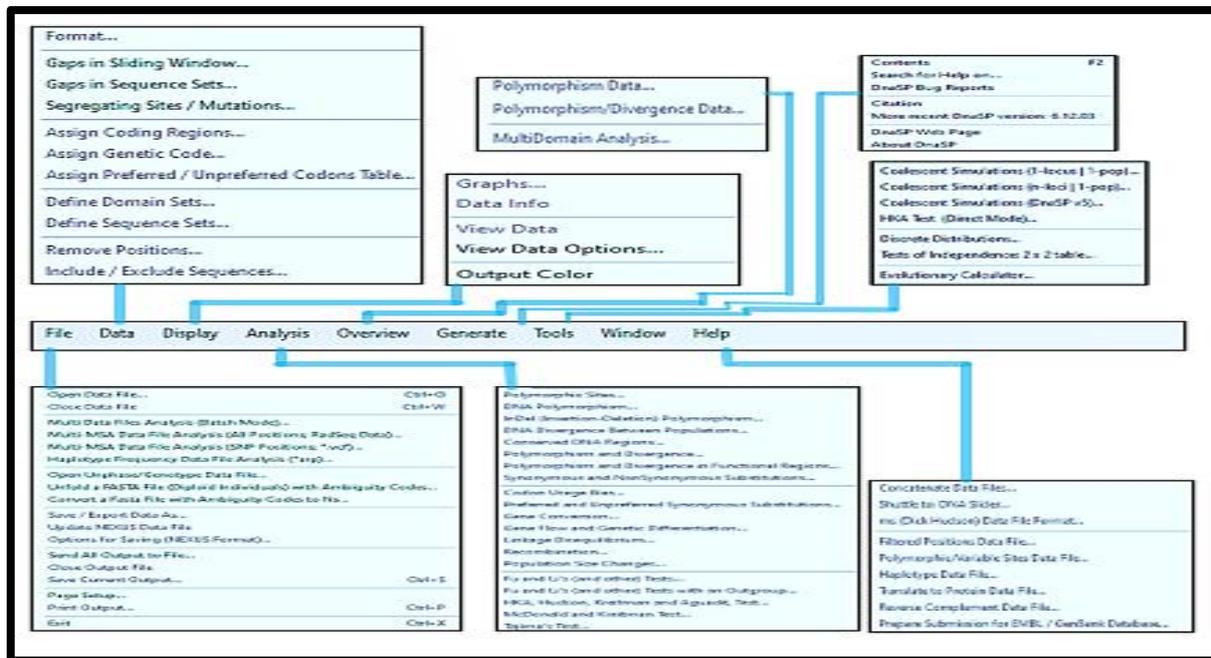


Figure 24 : Barre d'outils de DnaSP.

- Le DnaSP produit des fichiers de données sur les haplotypes. Les résultats peuvent être enregistrés sur des fichiers NEXUS ou de données de Roehl (**Figure25**).

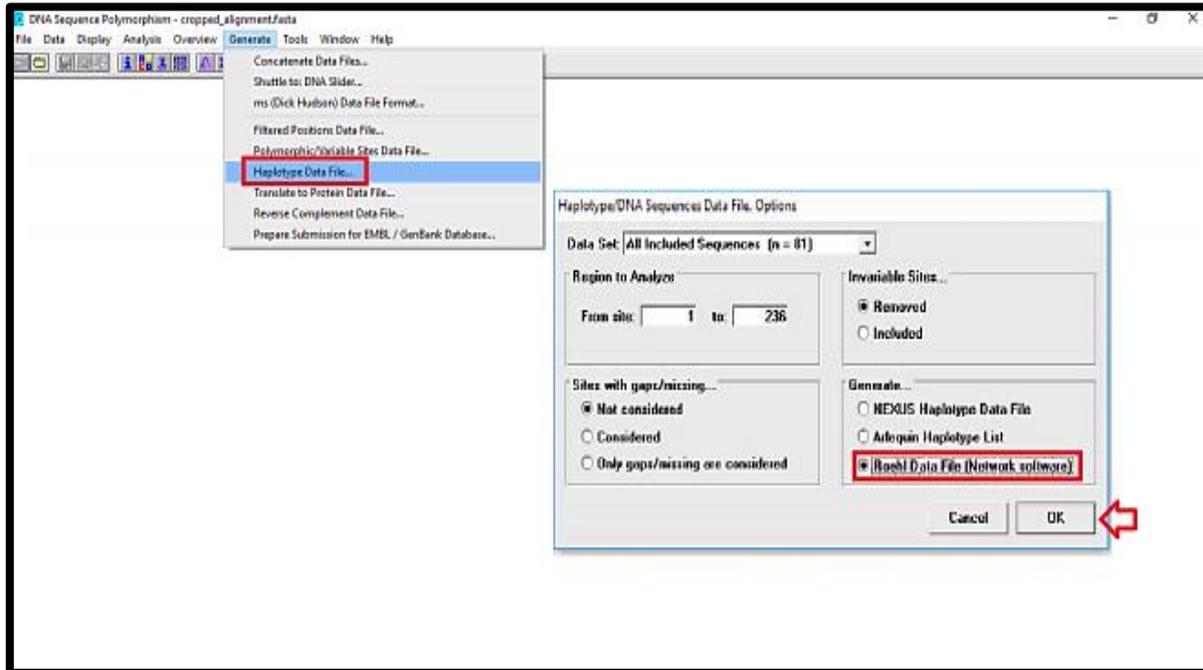


Figure 25 : Interface utilisateur de DnaSP.

II.2.1.1.4. Dambe :

C'est un logiciel qui intègre des paquets pour manipulation et analyse des séquences comportant une interface conviviale et un grand choix de fonctions analytiques de bioinformatique, phylogénétiques, et génomique descriptive et comparative (SALEMI et al., 2003; FELSENSTEIN, 2004 ; LEMEY et al., 2009). Permet de convertir, manœuvrant, statistiquement et graphiquement décrivant (**Figure26**).

- DAMBE peut représenter graphiquement le taux de C/G et A/T (**Figure 27**).

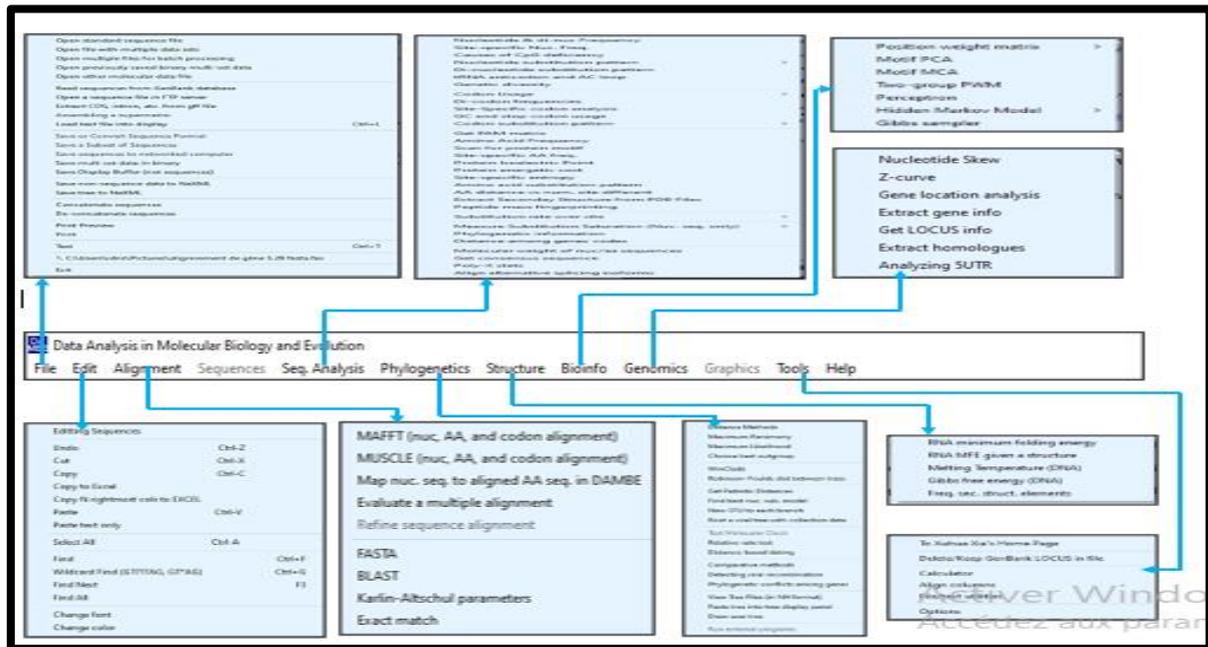


Figure 26 : Barre d'outil du programme DAMBE.

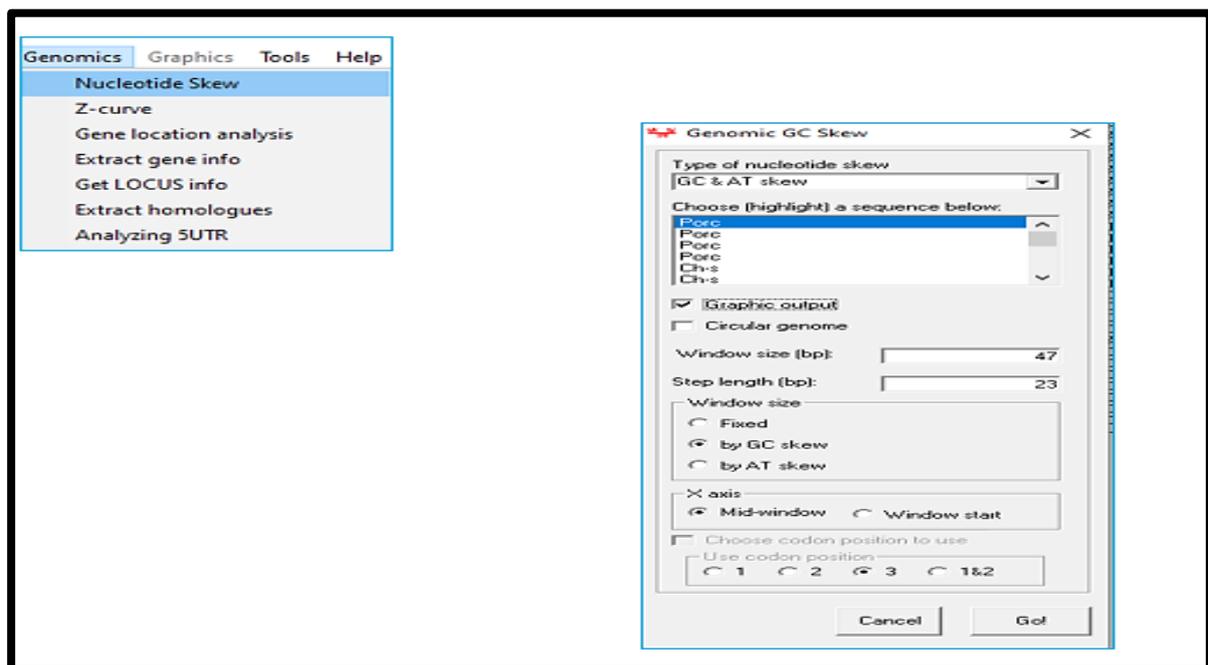


Figure 27 : Barre d'outils qui fait les graphes de C/G et A/T.

II.2.1.1.5. Network :

On a terminé notre travaille par le logiciel NETWORK 10.2(figure 28), est recommandé pour les présentations des séquences d'ADN ou d'acides aminés sous forme de réseau.

Network est recommandé pour : les présentations graphiques dans des publications de qualité, plus les présentations des diagrammes et secteurs, l'affichage de noms de nœuds plus longs (15 caractères) et de noms de lien (9 caractères), les nœuds proxy de réseau, les mutations par lien et les individus par nœud.

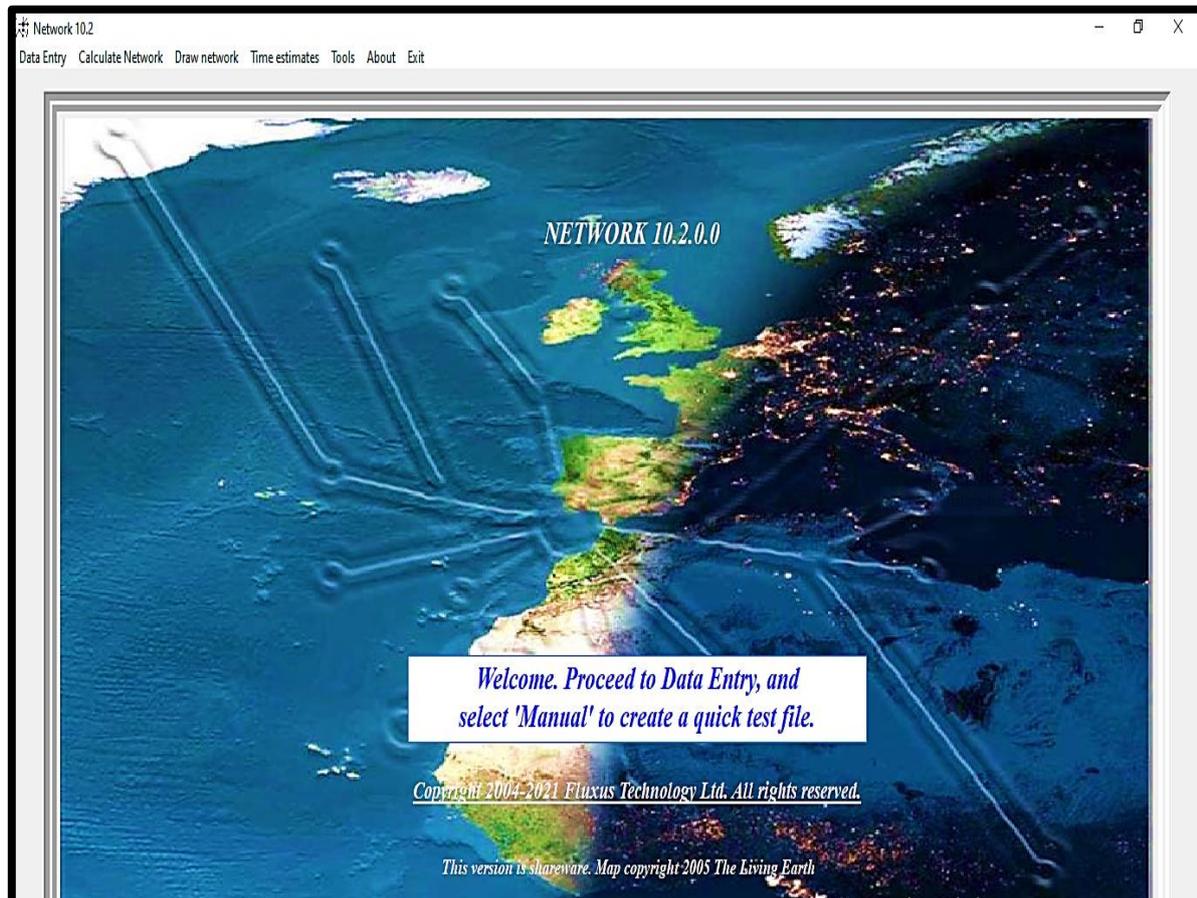


Figure 28 : La page d'accueil de Network 10.2.

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.2.1.1.5.1. Méthode 1 base sur les haplotypes

Pour le calcul du réseau initial, nous passons par différentes étapes.

1. Calculate Network menu.
2. Network calculations.
3. Median Joining.
4. File.

5. Open.
6. DNA data file (rdf) (**Figure29**).

Sélectionnez le dossier de rdf dont vous avez manuellement créé dans le rédacteur de données de Network ou exporté du programme DnaSP.

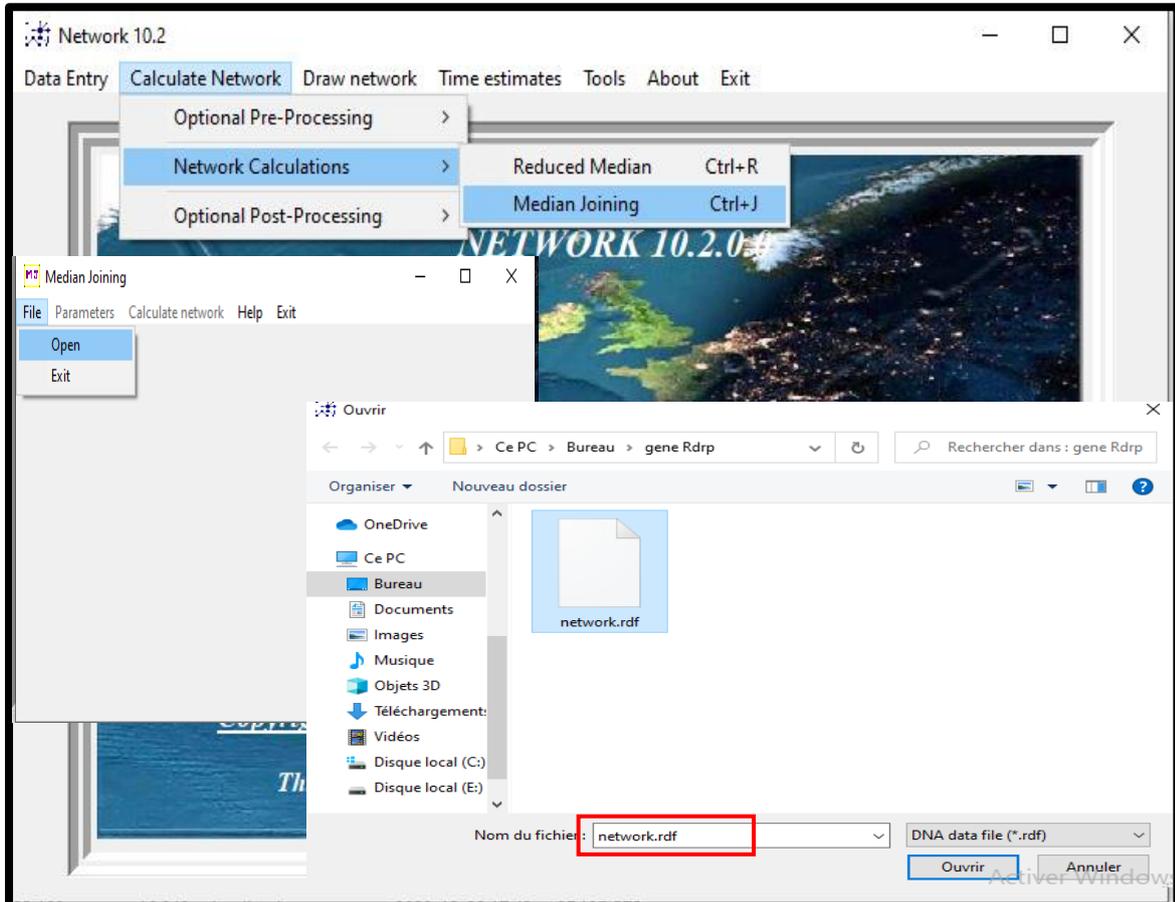


Figure 29 : Ouverture du fichier de données de rdf pour le calcul de MJ Network.

L'enregistrement de notre fichier peut être sous format *.out pour la reconstruction de réseau de NETWORK (**figure 30**).

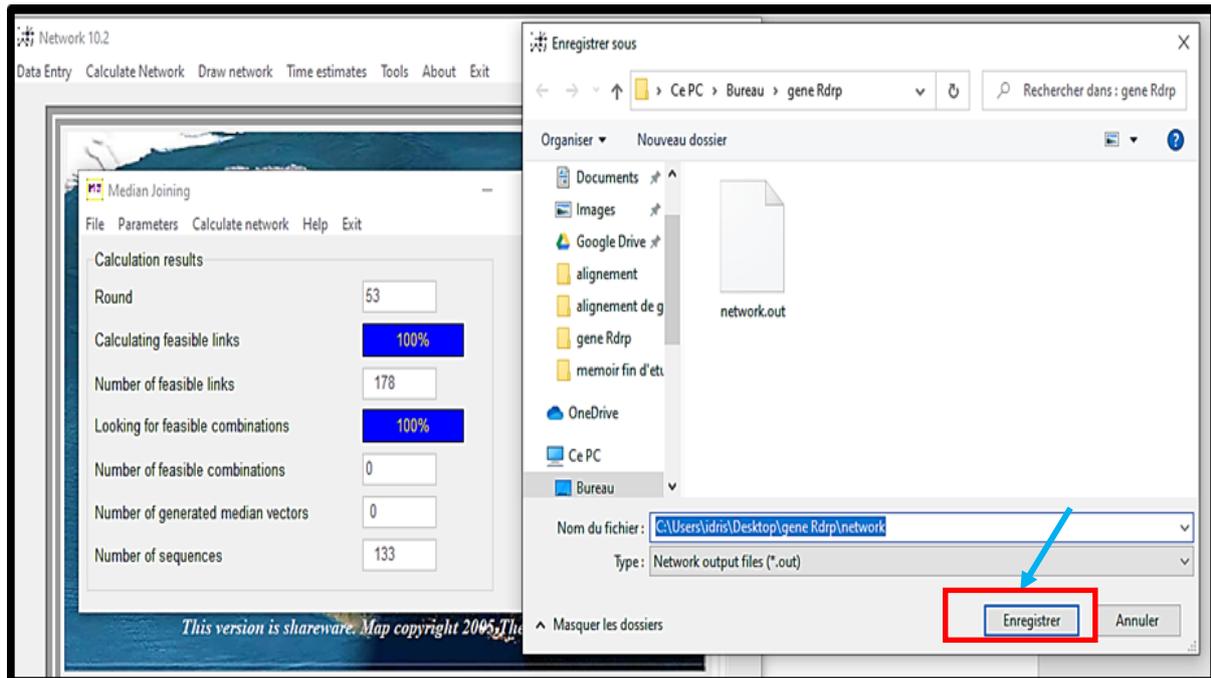


Figure 30 : Un fichier sous forme *.out a été sauvegardé.

Cliquer sur Draw network et entrer le fichier précédemment enregistré, pour obtenir le réseau haplotypique, puis l'enregistrer sous forme (**Figure 31**).

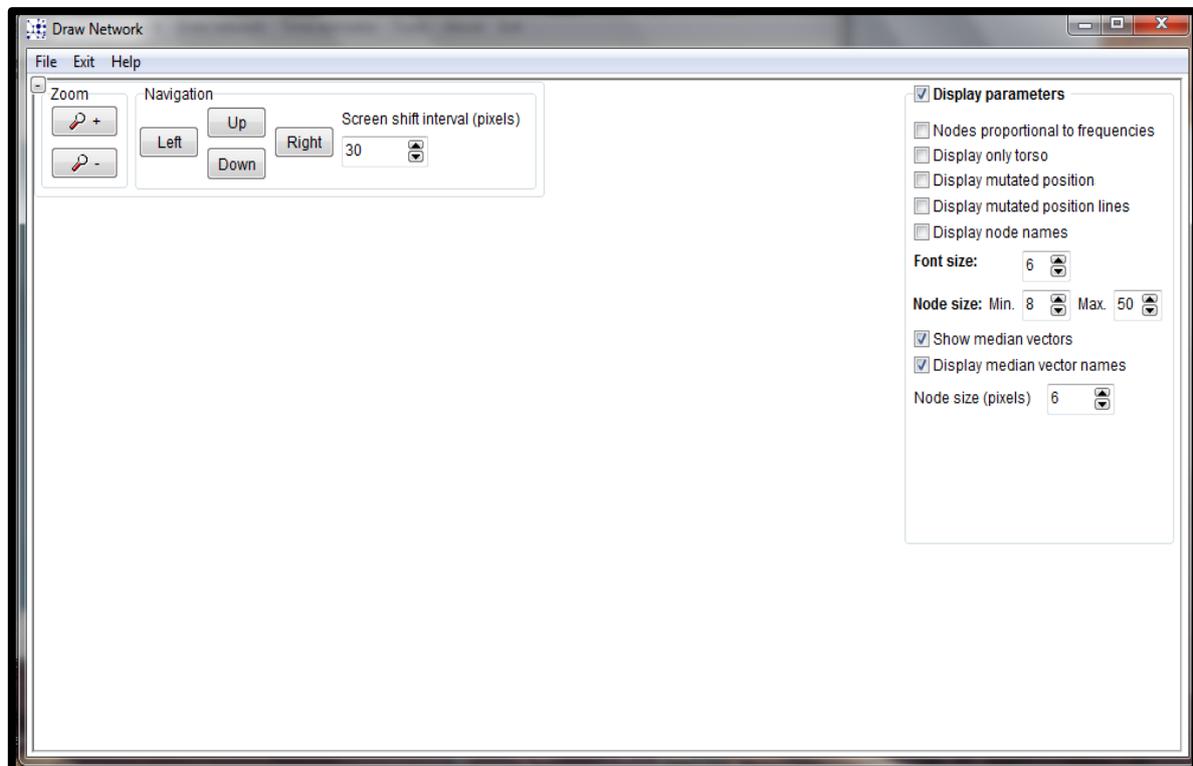


Figure 31 : Reconstruction de réseau haplotypique.

II.2.1.1.5. 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 RD

- 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 28 puis cliquer sur Convert to RDF.

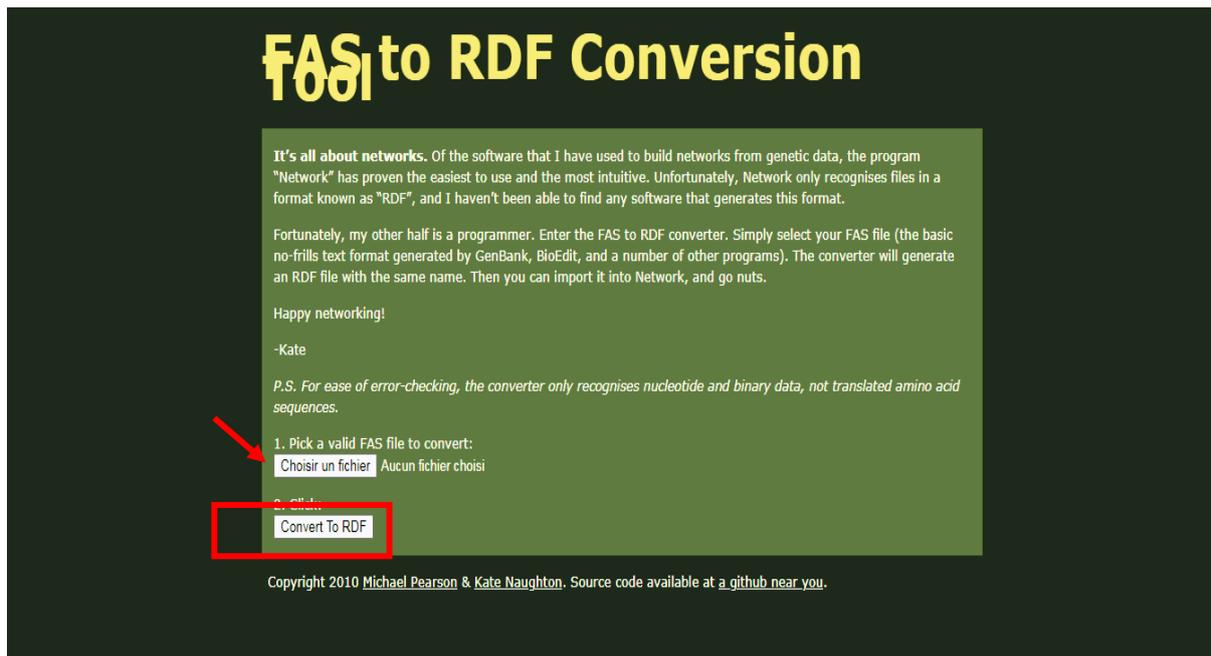


Figure 32 : Page d'accueil en ligne de programme FAS to RDF Conversion.

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 entre les séquences d'ADN de différentes souches virales du coronavirus publié 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 des positions des mutations sur les gènes analysés et d'avoir en même temps la composition nucléotidique de chaque gène grâce à des analyses statistiques.

III.1. Résultats

III.1.1. GenBank

Les séquences d'ADN de différentes souches virales du coronavirus extraites selon la taille et les espèces animales touchés à travers plusieurs pays dans le monde, sont mentionnées dans le tableau en annexe.

Après l'analyse des données de la GenBank nous avons récupéré plusieurs séquences de différentes souches de coronavirus mais nous n'avons choisi que les séquences nucléotidiques qui présentent deux gènes, gène de pointe (gène structurale) et gène RdRp (gène non structurale) qui sont en commun entre plusieurs espèces animales touchés (chauve-souris, rat, souris, porc ...) (Résumé dans le tableau III ci-dessous) par rapport aux autres gènes.

Tableau III : Séquences nucléotidique de différentes souches virales de coronavirus extraites de la GenBank.

Animal	Espèce	Région	Gène	Souche Virale
Porc	<i>Porcine</i>	Canada, USA, Chine, Thaïlande, Vietnam, Espagne	S	PRCV, PEDV PDCOV, TGEV, PHEV
	<i>Cochon</i>	Chine, Mexique, Corée du Sud		
	<i>Sus scrofa</i>	Mexique		
	<i>Cochon de nourrice</i>	Chine		
	<i>Pipistrellus kuhlii</i>	Italie		
	<i>Swine</i>			
Civette	<i>Civette palmiste</i>	Chine	S	Sars-CoV1
	<i>Chats civette palmiers</i>			

RESULTATS ET DISCUSSIONS

CHAPITRE III

Humain	<i>Homo Sapiens</i>	Chine, Arabie Saoudite, France, Pays-Bas, USA, Iran, Argentine, Slovénie	S et RdRp	HCoV-HKU1 Sars-CoV2 MERS-CoV HCoV-OC43, HCoV 229E	
Chameau	<i>Camelus dromedarius</i>	Maroc, Éthiopie	S	DcCoV	
Bovin	-	Corée du Sud	S	BCoV	
	-	Chine			
	-	Slovénie			RdRp
Ferret	<i>Blaireau furet chinois</i>	Chine	S	FeCoV	
Dinde	<i>Meleagris gallopavo</i>	USA	S	TCoV	
Pintade	<i>Pintade</i>	France		gfCoV	
Chat	<i>Felis catus</i>	Chine		RdRp	FCoV, FIP
	<i>Chat domestique</i>				
Chien	<i>Canis lupus familiaris</i>	Chine	CCoV		
Chauve-souris	<i>Tylonycteris pachypus</i>	Chine	RdRp		BatCoV
	<i>Lasiopodomys gregalis</i>				
	<i>Rhinolophus ferrumequinum</i>	Corée du Sud			
	<i>Ptenochirus jagori</i>	Philippines			
	<i>Eptesicus isabellinus</i>	Espagne			
	<i>Hipposideros sp</i>	Zimbabwe			
	<i>Taphozous perforatus</i>	Arabie Saoudite			
	<i>Eidolon helvum</i>				
	<i>Cynopterus sp</i>	Inde			
	<i>Nycteris thebaica</i>	Mozambique			
	<i>Mormopterus jugularis</i>	Madagascar			
	<i>Rhinolophus affinis</i>	Japon			
	<i>Eidolon helvum</i>	Kenya			
<i>Molossus molossus</i>	Brésil				

	<i>Molossus rufus</i>			
Poulet	<i>Poulet</i>	Chine	S	PDCOV
Rat	<i>Bandicota savilei</i>	Laos	S	RtBs -CoV
	<i>Apodemus draco</i>	Chine	RdRp	
	<i>Rattus flavipectus</i>	Chine		
	<i>Rattus nitidus</i>			
	<i>Niviventre</i>			
Yak	<i>Yak</i>	Chine	S	YakCoV
Canard	<i>Anas platyrhynchos</i>	POLOGNE	RdRp	ACoV
	<i>ANAS SP</i>			
Souris	<i>Mus carolis</i>		RdRp	Rt-CoV
	<i>Myodes rutilus</i>			
	<i>Myodes rufocanus</i>			
	<i>Apodemus sylvaticus</i>			

Afin d'établir les relations phylogénétiques entre les différentes souches de coronavirus, notre étude est basée sur deux gènes :

Gène S : Est la protéine avec les séquences les plus variables dans les génomes des coronavirus, la glycoprotéine S est caractérisée comme le déterminant essentiel de l'entrée virale dans les cellules hôtes qui se compose de deux sous-unités fonctionnelles, à savoir S1 et S2. La sous-unité S1 reconnaît et se lie au récepteur hôte via le domaine de liaison au récepteur (RBD) tandis que S2 est responsable de la fusion avec la membrane de la cellule hôte (Wrapp et al., 2020 ; Walls et al., 2020 ; Chen et al., 2020)

Gène RdRp : L'ARN polymérase dépendante de l'ARN (RdRp), également connue sous le nom de nsp12, est un composant central du complexe de réplication et de transcription du virus et gère la réplication et la transcription de l'ARN viral Nsp12 possède au moins deux domaines, le « domaine conservé à l'échelle du nidovirus avec activité nucléotidyl transférase (Lehmann et al., 2015) et le domaine canonique RdRp C-terminal (Gorbalenya et al., 1989). La séquence codant pour nsp12 contient le site de décalage de cadre ribosomique ORF1a/1b, et un événement de décalage de cadre -1 programmé entraîne la traduction de ORF1b pour produire la polyprotéine pp1ab qui contient nsp12. Le nsp12-RdRp des CoV est une cible

médicamenteuse principale, qui peut être inhibée dans la cellule hôte sans aucun effet secondaire toxique

Dans les coronavirus, la protéine de pointe est fonctionnellement associée à la reconnaissance des hôtes et l'ARN polymérase dépendante de l'ARN (RdRp) est lié à la prolifération du virus. (Qin *et al.*, 2015)

III.1.2. Construction des arbres phylogénétiques par MEGA

Afin de faire l'alignement pour construire les deux arbres phylogénétiques nous avons choisi les séquences représentatives de 2 gènes. Il y a plusieurs séquences présentant la même souche et le même pays mais ils sont de tailles (paires de bases nucléotidiques pb) différentes, pour cela nous avons choisi la séquence qui a la plus grande taille.

III.1.2.1. Arbre phylogénétique de gène de pointe (S)

L'arbre phylogénétique (**Figure 33**) a été réalisé sur la base du gène S de différentes souches virales du coronavirus qui ont touché l'espèce humaine et plusieurs espèces animales avec une taille qui varie entre (4374pb et 1992pb). Elle a été divisée en quatre groupes : Alpha CoV, Delta CoV, Gamma CoV, Beta CoV.

Alpha CoV (Alphacoronavirus) :

Aa (α): Il comprend les séquences nucléotidiques de la gastro-entérite transmissible (*TGEV*) de la Chine (4344pb) ; le coronavirus respiratoire porcine (*PRCV*) de l'Espagne (3678pb) qui infecte le porc (porcine, cochon), le coronavirus Canine (*CCoV*) de la Chine (4374pb) infecte le chien (*Canis lupus familiaris*), le coronavirus *Félin CoV* de la Chine infecte le chat (*Felis catus*, chat domestique), le coronavirus de Ferret (*FRCoV*) de la Chine (4374pb) infecte le ferret (*Melogale moschata*).

Ab (α b) : Il représente les séquences de souche virale la diarrhée épidémique porcine (*PEDV*) du Canada (4271pb), Mexique (4200pb), Chine, Vietnam, Thaïlande, Japon (4161pb). Elle infecte les porcs (*sus scrofa*, cochon).

Delta CoV (Deltacoronavirus) : Il représente la séquence de souche Delta CoV de porc (*PDCoV*) des Etats- unis, Mexique (3483pb), et la Chine (3480pb), elle infecte les porcs (cochon de nourrice, porcine, *Sus scrofa*) et *PDCoV* de chine (3480pb) infecte le poulet.

Gamma CoV (Gamma coronavirus γ) : Il comprend les séquences de coronavirus de la Turquie TCoV, d'USA (3627 pb) infecte la dinde (*Meleagris gallopavo*) et le coronavirus de pintade gfCoV de France (3591pb).

Beta CoV (Betacoronavirus) :

Ba (β a) : Il regroupe les souches du coronavirus du syndrome respiratoire aiguë sévère SARS-CoV1 de la Chine (3768pb) infecte la civette (civette palmiste, chats civette palmiers) ; le coronavirus du syndrome respiratoire aiguë sévère SARS-CoV2 de USA et Iran (3822pb) infecte l'humain (*Homo sapiens*).

Bb (β b) : Il comprend le coronavirus lié au syndrome respiratoire de Moyen-Orient (MERS-CoV) de Saudi Arabia (4062pb) infecte l'humain (*Homo sapiens*), le coronavirus de chauve-souris (*Bat-CoV*) de la Chine (4059 pb) infecte les chauves-souris (*Tylonycteris pachypus*).

Bc (β c) : Il comprend le coronavirus humain *HCoV-HKU1* des Pays-Bas, la Chine (4071pb) infecte l'humain (*Homo sapiens*).

Bd (β d) : Il comprend le coronavirus dromadaire DcCoV d'Ethiopie (4101pb) infecte le chameau *Camelus dromedarius*, le coronavirus humain HCoV-OC43 de la France (4077pb), le coronavirus de bovin BCoV de la Coré du Sud, la Chine (4092pb) infecte bovin (*sage, veau*) ; le coronavirus Yak de la Chine (4092pb).

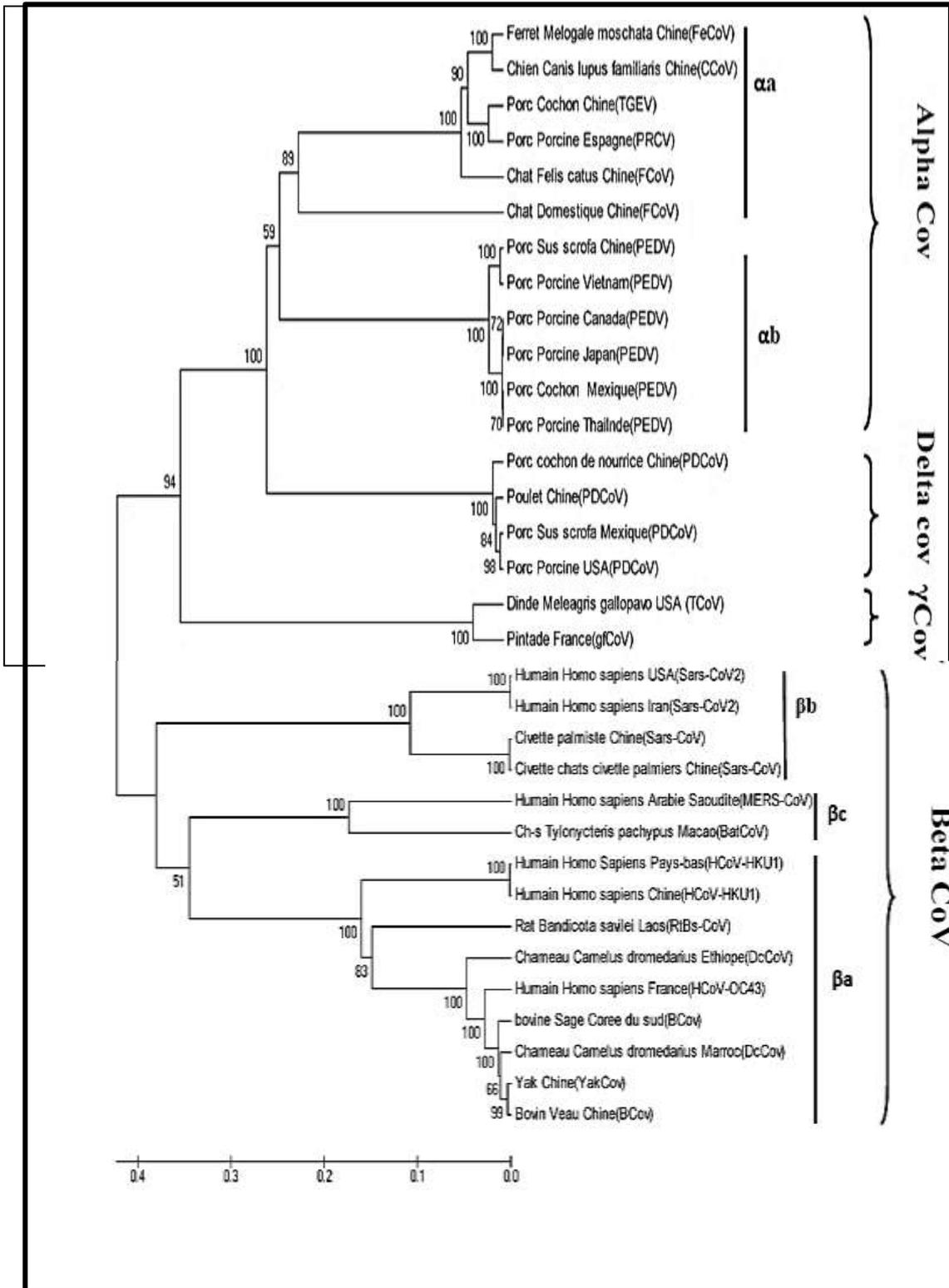


Figure 33 : Arbre phylogénétique de gène S de différentes souches de coronavirus construit à partir 33 séquences en utilisant la méthode UPGMA par logiciel MEGA5.

III.1.2.2. Arbre phylogénétique de gène RdRp

L'arbre phylogénétique (**Figure 34**) a été réalisé sur la base du gène RdRp de différentes souches virales de coronavirus qui ont touché l'homme et plusieurs espèces animales avec une taille qui varie entre (699pb et 251pb). Elle été divisé en : les Coronavirus non classés, bêtacoronavirus, gammacoronavirus, alphacoronavirus.

Coronavirus non classés

Après une recherche sur BLAST nous avons vu que les coronavirus non classés appartiennent aux groupes suivants :

- La séquence de ch-s (*T perforatus*) a une similarité de 98,99% avec une séquence de bovin de bêtacoronavirus par un code d'accession (AF124985).

-Les séquences de rat (*Lasiopodomys gregalis*), (*Rattus flavipectus*), (*Apodemus draco*), (*Mus carolis*), (*Nivivnetaire Niviventaire*), (*Apodemus sylvaticus*), (*Myodes refocanus*) des pays d'Asie ont une similarité de 97,71% avec une séquence de souris de bêtacoronavirus par un code d'accession (MT222256).

-La séquence de ch-s (*Necterus thebiaca*) de Mozambique a une similarité de 91,71% avec une séquence de ch-s (*Miniopterus sp*) de bêtacoronavirus par un code d'accession (MK569443).

-Les séquences de ch-s (*Edilum helvum*) de kenya et d'UKA ont une similarité de 98,43% avec une séquence de ch-s (*Rousettus*) de bêtacoronavirus par un code d'accession (MG693171).

-La séquence de ch-s (*Mormopterus jugularis*) de Madagascar a une similarité de 89,65% avec une séquence de ch-s (*Chéréphon sp*) d'alphacoronavirus par un code d'accession (HQ728486).

-La séquence de ch-s (*Molessus Molessus*) du Brésil a 98,70% d'identité avec une séquence de ch-s (*M rufus*) d'alphacoronavirus par un code d'accession (KC886321).

-La séquence de ch-s (*Molessus retilus*) du Brésil a 84,28% d'identité avec une séquence de rat (*Rattus norvidus*) d'alphacoronavirus par un code d'accession (MW011453).

-La séquence de ch-s (*Rinolophus affinis*) du Japon a une identité de 94,56% avec une séquence de porc d'alphacoronavirus par un code d'accession (MT039232).

-La séquence de souris (*Myodes retilus*) de Chine a une identité de 93,41% avec une séquence de rat (*Rattus norvigicus*) d'alphacoronavirus par un code d'accèsion (MW011455).

-La séquence de rat (*Rattus nitidus*) de Chine a une identité de 97,69% avec une séquence de rat (*Rattus norvegicus*) d'alphacoronavirus par un code d'accèsion (MW011455).

Bétacoronavirus

Les béta coronavirus dans cet arbre comprennent les séquences suivantes :

βa: Renferme le coronavirus BtCov il infecte les chauves-souris de l'Arabie Saoudite de (398pb) et de (400pb), et des Philippines (393pb). Comprennent aussi la séquence du coronavirus de bovin BCoV de la Slovénie (390pb), le coronavirus de l'Encéphalomyélite Hémagglutinante Porcine PHEV (396pb), le coronavirus de rat RtCoV de (440pb), (426pb), (407pb), (428pb) de la Chine et une séquence du Vietnam (407pb).

βb: Ce sous genre renferme la souche BtCov infectant la ch-s du Mozambique (442pb).

βc: Ce sous genre renferme la souche de coronavirus ECoV qui infecte le hérisson de la Chine (562pb), le coronavirus de souris RtCoV de la Chine (infectant aussi les rats) de taille (440pb), (428pb), (439pb), la souche BtCov de l'Espagne (396pb), de la Corée du Sud (603pb), du Kenya (448pb), l'Inde (401pb),

Béthacoronavirus non classé

Après une autre recherche sur blast nous avons :

La séquence de ch-s (*T perforatus*) de l'Arabie Saoudite révèle une similitude 99.25 % avec une séquence du bovin qui fait partie du sous genre Embicovirus (**a**) appelé aussi **2a** avec un code d'accèsion (KY502388) et avec un même code d'accèsion, une similitude de 86.28 % avec la séquence de souris (*Apodemus Sylavatecus*) de la chine.

La séquence de rat (*Lasipodemus gregalis*) de la Chine à un pourcentage de similarité de 91.84 avec une séquence d'*Homo Sapiens* qui fait partie au sous genre **a** avec un code d'accèsion (KU745556).

Par un code d'accèsion (MT820630) et une similitude de 96.56 %, la séquence de rat du Vietnam est proche d'une autre séquence de rat qui fait partie du sous genre **a**.

RESULTATS ET DISCUSSIONS

CHAPITRE III

Avec une similarité de 95.92 %, la séquence de rat (*Rattus flavipectus*) de la Chine est proche de la séquence de souris classé dans le sous genre **a** par un code d'accension (MT221701).

La séquence de la souris (*Mus carolis*) de la chine a une similitude de 96.69 % avec une séquence de murine par un code d'accension (MT221985) fait partie du sous genre **a**.

La séquence de souris de la Chine (*Myodes rufocanus*) possède une similitude de 80.49 % avec une séquence d'Homo sapiens classée au sous genre **a** par code d'accension (KU745556).

La séquence ch-s (*Eptesicus isabellinus*) de l'Espagne montre une similitude de 94.95 % par un code d'accension (MG021452) avec la séquence du hérisson (*Ericanus amurensis*) de la Chine qui fait partie du sous genre Merbicovirus (**b**).

La séquence de ch-s (*Nycteris thebiaca*) du Mozambique possède un pourcentage de similitude de 75.82 avec une séquence classé au sous genre de Sarbicovirus (**c**) par un code d'accension (AB889999).

Gammacoronavirus

Ce groupe renferme l'ACov (Avian Coronavirus) qui infecte le canard (*Anas platyrhynchos*) de Pologne (684pb), et (*Anas sp*) (699pb) du même pays et même animal.

Alphacoronavirus

Renferme le (*FRSCV*) qui infecte le ferret avec une taille de séquence de (393pb) du Pérou, le (*HCoV-229E*) de l'humain avec une taille de séquence de (404pb) du Brésil et de Slovénie (390pb), le coronavirus (*BtCoV*) de Zimbabwe (415pb), Madagascar (442pb), Brésil (426pb) et (404pb), Japon (440pb), le (*RtCoV*) infectant les souris et les rat (440pb) (433pb) de la Chine.

Remarque : Pour le gène RdRp nous n'avons pas mais les sous-groupes dans le clade d'Alphacov car nous n'avons pas trouvé des séquences similaires par Blast sur la GenBank.

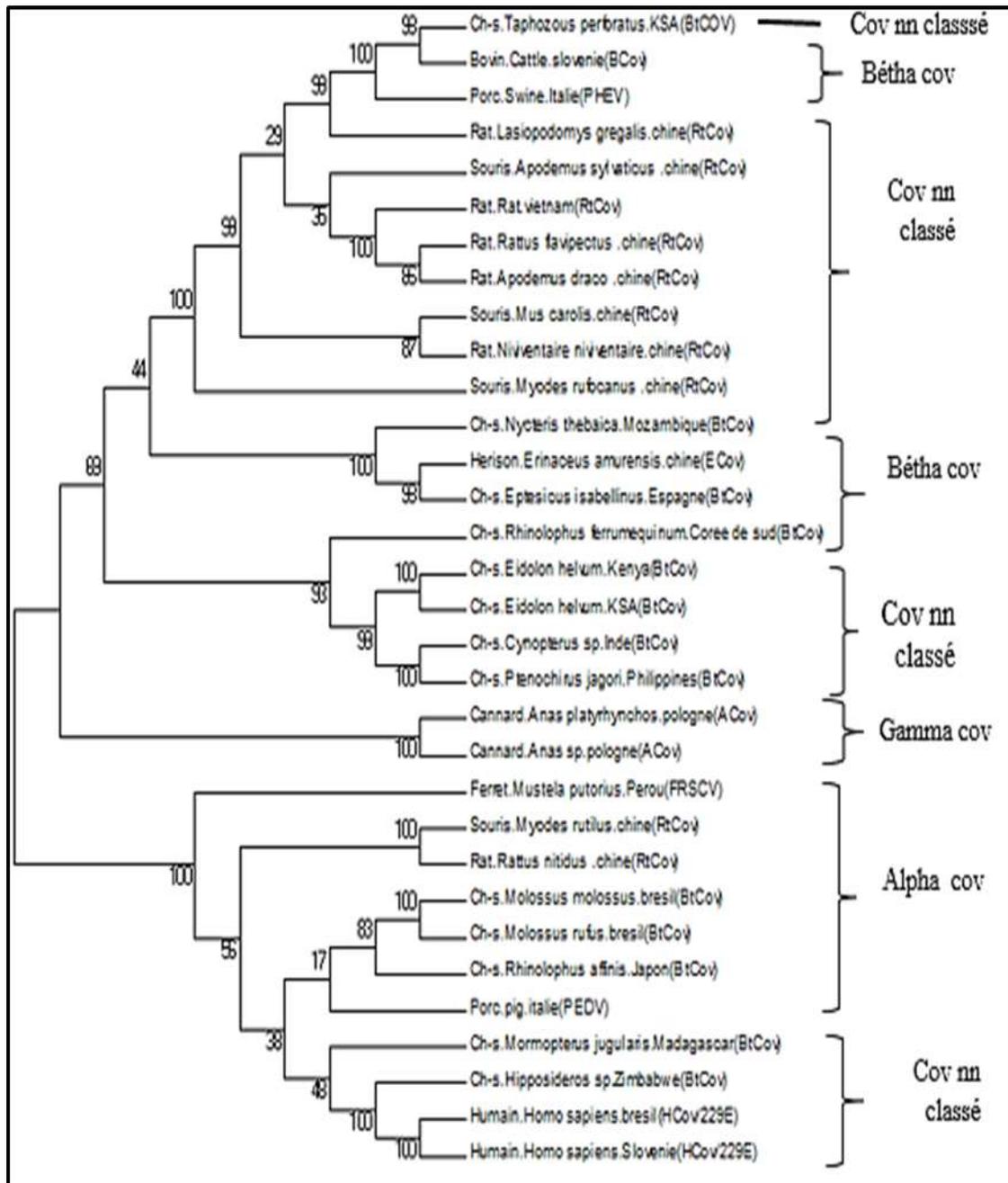


Figure 34 : Arbre phylogénétique construit par la méthode UPGMA basé sur 32 séquences de gène Rrdp des souches virales de coronavirus en utilisant le logiciel MEGA5.

- ❖ Après des alignements avec BLAST nous avons pu classer les souches virales dans l'arbre phylogénétique (**Figure 35**) comme suit :

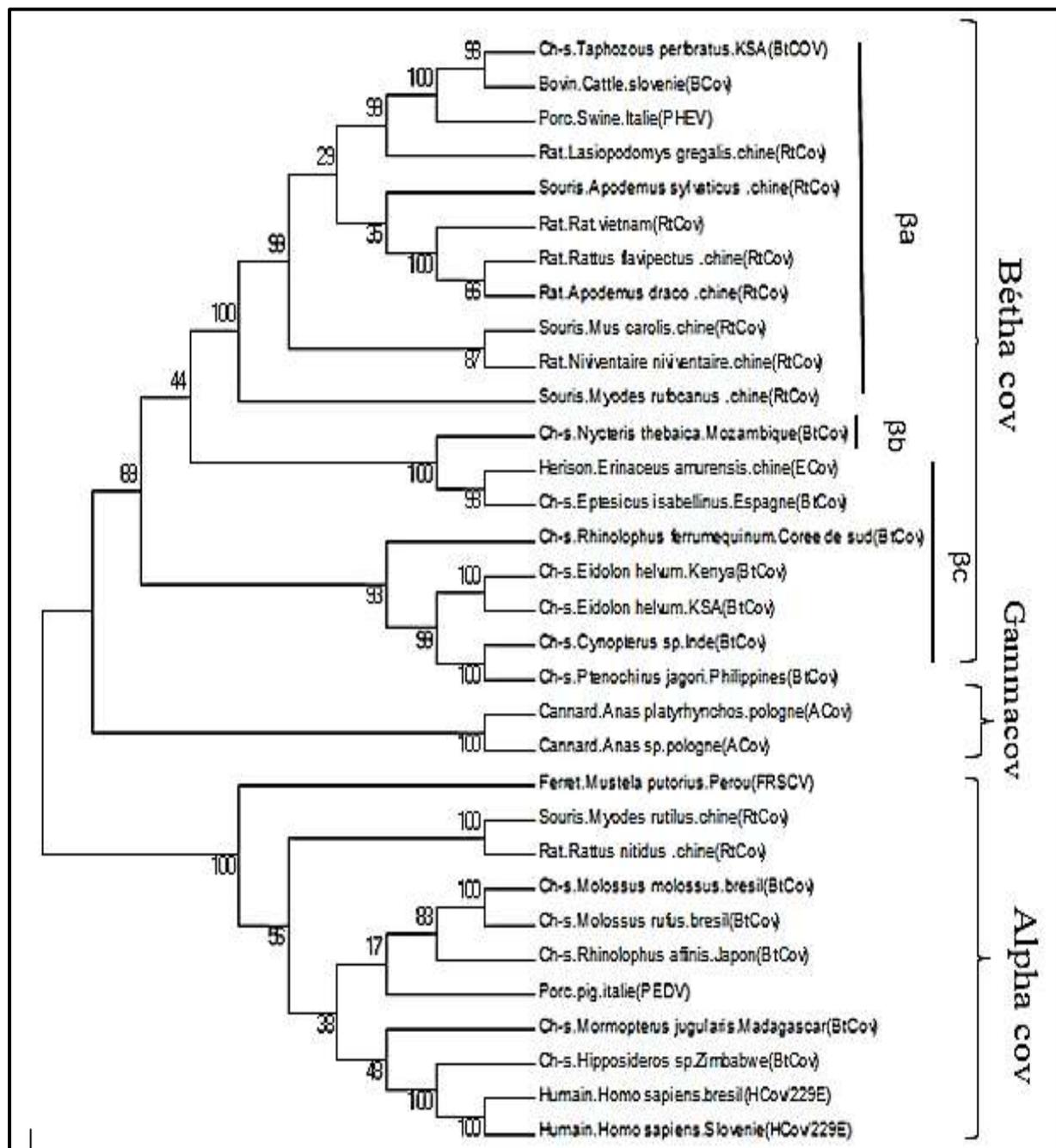


Figure 35 : L'arbre phylogénétique final construit par la méthode UPGMA basé sur 32 séquences de gène RdRp des souches de coronavirus utilisant le logiciel MEGA5.

III.1.3. DnaSP

L'analyse des gènes S et RdRp par le logiciel DnaSP v.6.12.03 révèle une diversité génétique très importante, un total de 1434 sites polymorphes et 31 haplotypes ont été trouvés dans les 33 séquences étudiées, avec une diversité haplotypique (hd) significative positive égale à 0.996

RESULTATS ET DISCUSSIONS

CHAPITRE III

pour le gène S (**Figure 36**) et un total de 858 sites polymorphes et 34 haplotypes ont été trouvé dans 32 séquences étudiées avec une diversité haplotypique (hd) de 0,9983 pour le gène Rdrp. (**Figure 37**).

```
Output of: cropped_alignment (11).fasta
Haplotype/DNA Sequences Data File

Input Data File: C:\...\cropped_alignment (11).fasta
Number of sequences: 33   Number of sequences used: 33
Selected region: 1-1247   Number of sites: 1247
Total number of sites (excluding sites with gaps / missing data): 1058

Sites with alignment gaps: not considered
Number of variable sites: 862

===== Haplotype Distribution =====
Number of haplotypes, h: 31
Haplotype diversity, Hd: 0,9962

Hap_1: 1 [1]
Hap_2: 1 [2]
Hap_3: 1 [3]
Hap_4: 1 [4]
Hap_5: 2 [5 7]
Hap_6: 2 [6 9]
Hap_7: 1 [8]
Hap_8: 1 [10]
Hap_9: 1 [11]
Hap_10: 1 [12]
Hap_11: 1 [13]
Hap_12: 1 [14]
Hap_13: 1 [15]
Hap_14: 1 [16]
Hap_15: 1 [17]
Hap_16: 1 [18]
Hap_17: 1 [19]
Hap_18: 1 [20]
Hap_19: 1 [21]
Hap_20: 1 [22]
Hap_21: 1 [23]
Hap_22: 1 [24]
Hap_23: 1 [25]
Hap_24: 1 [26]
Hap_25: 1 [27]
Hap_26: 1 [28]
Hap_27: 1 [29]
Hap_28: 1 [30]
Hap_29: 1 [31]
Hap_30: 1 [32]
Hap_31: 1 [33]

Hap_1: 1 [Ferret]
Hap_2: 1 [Chien]
Hap_3: 1 [Chat]
Hap_4: 1 [Porc]
Hap_5: 2 [Porc Porc]
Hap_6: 2 [Porc Porc]
Hap_7: 1 [Porc]
Hap_8: 1 [Porc]
Hap_9: 1 [Chameau]
Hap_10: 1 [bovine]
Hap_11: 1 [Yak]
Hap_12: 1 [Bovine]
Hap_13: 1 [Chameau]
Hap_14: 1 [Humain]
Hap_15: 1 [Humain]
Hap_16: 1 [Humain]
Hap_17: 1 [Humain]
Hap_18: 1 [Ch-s]
Hap_19: 1 [Humain]
Hap_20: 1 [Humain]
Hap_21: 1 [Civette]
Hap_22: 1 [Civette]
Hap_23: 1 [Porc]
Hap_24: 1 [Dinde]
Hap_25: 1 [Pintade]
Hap_26: 1 [Rat]
Hap_27: 1 [Porc]
Hap_28: 1 [Porc]
Hap_29: 1 [Poulet]
Hap_30: 1 [Porc]
Hap_31: 1 [Chat]
```

Figure 36 : Nombre des haplotypes de gène S calculé par DnaSP.

```

Output of: result (1) fasta forma.fas
Haplotype/DNA Sequences Data File
Input Data File: C:\...\result (1) fasta forma.fas
Number of sequences: 35   Number of sequences used: 35
Selected region: 1-858   Number of sites: 858
Total number of sites (excluding sites with gaps / missing data): 210
Sites with alignment gaps: not considered
Number of variable sites: 202
===== Haplotype Distribution =====
Number of haplotypes, h: 34
Haplotype diversity, Hd: 0.9983
Hap_1: 1 [1]
Hap_2: 1 [2]
Hap_3: 1 [3]
Hap_4: 1 [4]
Hap_5: 1 [5]
Hap_6: 1 [6]
Hap_7: 1 [7]
Hap_8: 1 [8]
Hap_9: 1 [9]
Hap_10: 1 [10]
Hap_11: 1 [11]
Hap_12: 1 [12]
Hap_13: 1 [13]
Hap_14: 1 [14]
Hap_15: 1 [15]
Hap_16: 1 [16]
Hap_17: 1 [17]
Hap_18: 1 [18]
Hap_19: 1 [19]
Hap_20: 1 [20]
Hap_21: 1 [21]
Hap_22: 1 [22]
Hap_23: 2 [23 32]
Hap_24: 1 [24]
Hap_25: 1 [25]
Hap_26: 1 [26]
Hap_27: 1 [27]
Hap_28: 1 [28]
Hap_29: 1 [29]
Hap_30: 1 [30]
Hap_31: 1 [31]
Hap_32: 1 [33]
Hap_33: 1 [34]
Hap_34: 1 [35]

Hap_1: 1 [Porc.pig.italie]
Hap_2: 1 [Cannard.Anas]
Hap_3: 1 [Cannard.Anas]
Hap_4: 1 [Ch-s.Rhinolophus]
Hap_5: 1 [Herison.Erinaceus]
Hap_6: 1 [Ch-s.Eidolon]
Hap_7: 1 [Ch-s.Nycteris]
Hap_8: 1 [Ch-s.Mormopterus]
Hap_9: 1 [Ch-s.Rhinolophus]
Hap_10: 1 [Souris.Myodes]
Hap_11: 1 [Ch-s.Lasiopodomys]
Hap_12: 1 [Souris.Mus]
Hap_13: 1 [Rat.Niviventaire]
Hap_14: 1 [Rat.Rattus]
Hap_15: 1 [Souris.Myodes]
Hap_16: 1 [Rat.Rattus]
Hap_17: 1 [Souris.Apodemus]
Hap_18: 1 [Rat.Apodemus]
Hap_19: 1 [Ch-s.Molossus]
Hap_20: 1 [Ch-s.Hipposideros]
Hap_21: 1 [Rat.Rat.vietnam]
Hap_22: 1 [Ch-s.Molossus]
Hap_23: 2 [Humain.Homo Humain.Homo]
Hap_24: 1 [Ch-s.Cynopterus]
Hap_25: 1 [Ch-s.Eidolon]
Hap_26: 1 [Ch-s.Taphozous]
Hap_27: 1 [Porc.Swine.Italie]
Hap_28: 1 [Ch-s.Eptesicus]
Hap_29: 1 [Ferret.Ferret.Australie]
Hap_30: 1 [Ch-s.Ptenochirus]
Hap_31: 1 [Bovin.Cattle.slovenie]
    
```

Figure 37 : Nombre des haplotypes de RdRp calculé par DnaSP.

III.1.4. DAMBE

Le DAMBE est spécialisé dans les calculs et les statistiques moléculaires, les résultats de ce logiciel sont présentés sous 3 formes :

1. Les séquences identiques de chaque gène ; **la figure 38** est un exemple de gène S.
2. Calcul pour chaque gène (RdRp, S) : $(G-C)/(G+C)$, $(A-T)/(A+T)$, AT skews (la somme $(A-T)/(A+T)$) et GC skews (la somme $(G-C)/(G+C)$), le tableau IV présente le GC et le AT Skews pour les 2 gènes étudiées . **La figure 39** montre un exemple des tableaux donné par DAMBE.
3. La forme graphique de GC et AT pour les gènes S, RdRp ; un exemple de graphe de GC et AT skews de séquences de gène S de la souche virale FeCoV qui infecte le ferret (**Figure 40**).

Tableau IV : Les variations globales AT et GC skew des gènes S et RdRp.

Calcul Gènes	Global GC Skews		Global AT skews	
	Moins variable (Min)	Plus variable (Max)	Moins variable (Min)	Plus variable (Max)
Gène S	Ferret (FeCoV Chine)	Humain (SARS-2 USA)	Chameau (DcCoV Ethiopie)	Humain (MERS-CoV KSA)
AT/CG skews	0.5732	-0.04906	0.7022	-0.0514
Gène RdRp	Canard (ACoV)	Ch-s(BtCoV)	Ch-s(BtCoV)	Ch-s(BtCoV)
AT/CG skews	0.14286	-0.03175	0.17143	-0.11475

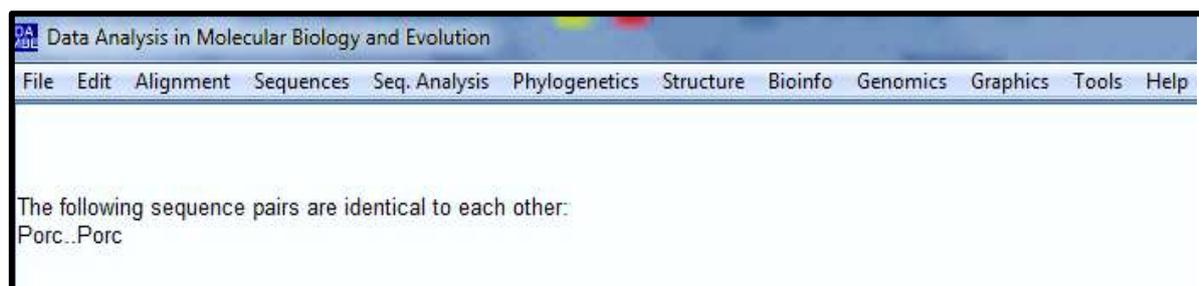


Figure 38 : Un exemple des séquences identiques de gène S calculé par DAMBE.

Nucleotide skews for sequence: Ferret
 Global GC Skew = 0,05526
 Global AT Skew = 0,04762
 Sliding window size optimized with GC skew.
 Window size = 161; stepsize = 7
 Circular genome: No

Window	Mid-window	(G-C)/(G+C)	(A-T)/(A+T)
0	80	-0,0417	0,0796
1	87	-0,0204	0,0714
2	94	-0,0769	0,1009
3	101	0,0370	0,1028
4	108	0,0741	0,1028
5	115	0,0357	0,1048
6	122	0,0690	0,0874
7	129	0,0357	0,0857
8	136	0,0189	0,1111
9	143	0,0385	0,1193
10	150	0,0000	0,1560
11	157	-0,0769	0,1193
12	164	-0,1154	0,1193
13	171	-0,1111	0,1215
14	178	-0,1111	0,1215
15	185	-0,0714	0,1238
16	192	-0,0714	0,1429
17	199	-0,1228	0,1346
18	206	-0,1379	0,1650
19	213	-0,1186	0,1176
20	220	-0,1579	0,1346
21	227	-0,1579	0,0769
22	234	-0,1034	0,0680
23	241	-0,1475	0,0800

Figure 39 : Les analyses de séquences de coronavirus du ferret donné par DAMBE.

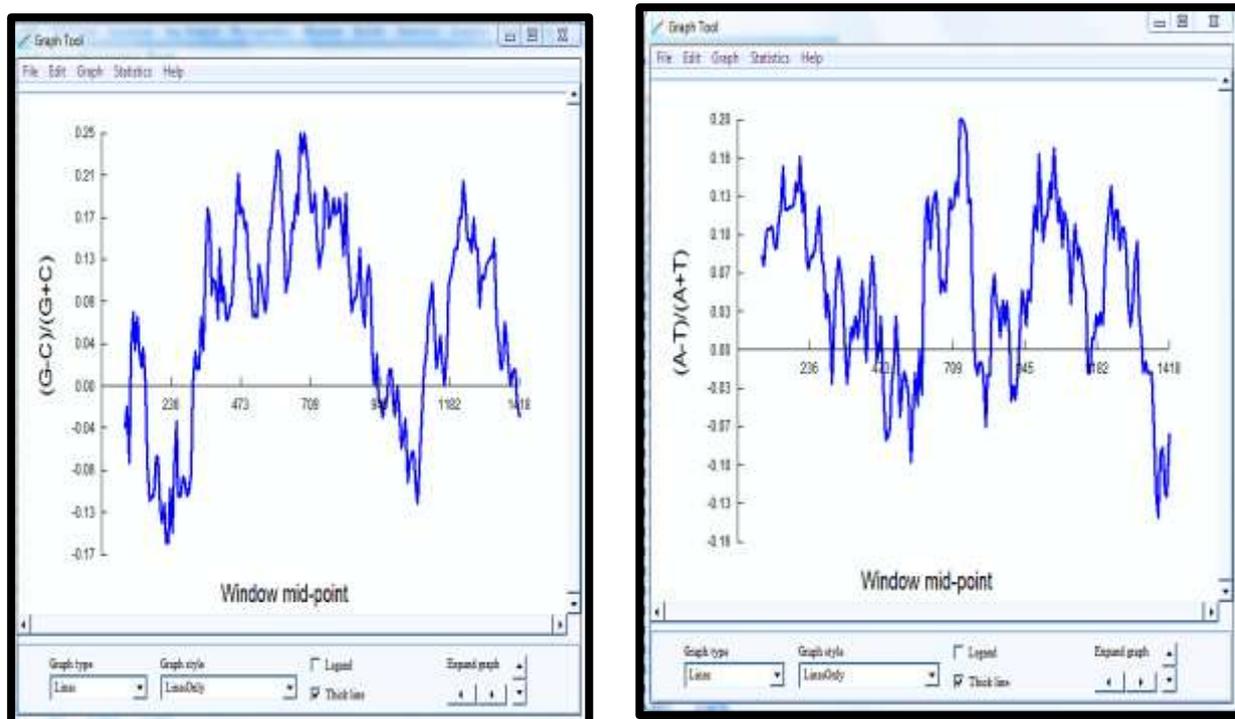


Figure 40 : La variation de GC et AT skews de gène S (séquence de coronavirus du ferret).

III.1.5. NetWork

Nous avons appliqué deux méthodes pour construire les réseaux phylogénétiques de gènes S et RdRp.

III.1.5.1 Résultat de la méthode 1 :

Nous nous sommes basés sur les haplotypes, nous avons utilisé les résultats de DnaSP pour obtenir les réseaux haplotypiques, différents réseaux d'haplotypes pour les gènes S (**Figure 41**) et RdRp (**Figure 42**).

Concernant le gène S nous avons obtenu quatre haplogroupes.

Haplogroupe 1 comprend les haplotypes suivants H1, H2, H3, H4, H7, H23 et H31 qui présentent une seule séquence avec couleur des cercles jaune à l'inverse des haplotypes H5 et H6 présenté par des couleurs différentes parce qu'il y a l'inclusion de deux séquences, H5 (séquence PEDV du Japon avec PEDV du Canada) et H6 (PEDV de Thaïlande avec PEDV du Mexique).

Haplogroupe 2 comprend les haplotypes suivants H27, H28, H29 et H30 chacune est présentée par une seule séquence (cercles jaune).

Haplogroupe 3 comprend les haplotypes H24 et H25, chacune est présentée par une seule séquence (cercles jaune).

Haplogroupe 4 comprend les haplotypes suivants : H9, H10, H11, H12, H13, H14, H15, H16, H17, H18, H19, H20, H21, H22, H24, H25 chacune est présentée par une seule séquence (cercles jaune).

Concernant le gène RdRp Nous avons obtenu :

Haplogroupe 1 comprend les haplotypes suivants : H7, H28, H5, H27, H11, H31, H26, H14, H21, H18, H13, H17, H12, H15, H30, H24, H26, H25 et H4 représenté tous par une séquence (cercles jaunes).

Haplogroupe 2 comprend les haplotypes H2 et H3 chacune est présentée par un cercle jaune (une seule séquence).

Haplogroupe 3 comprend les haplotypes : H10, H1, H16, H22, H19, H29, H9, H8, H20, H23 tous sont représentée par une seule séquence sauf le H23 il est représenté par deux séquences (le cercle divisée en noir et jaune).

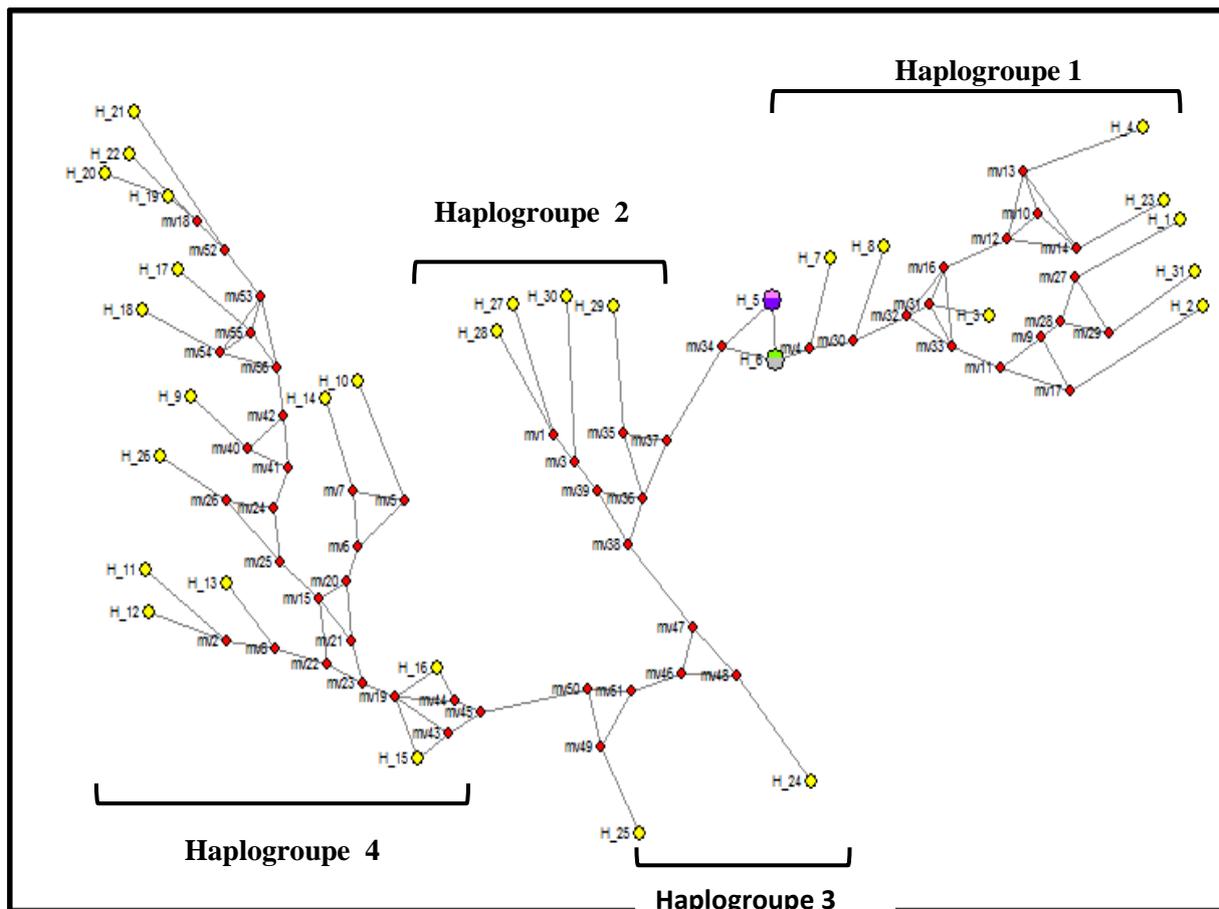


Figure 41: Réseau haplotypique des séquences de gène S réalisé par le logiciel NetWork basé sur les souches virales de coronavirus dans différentes régions géographiques (Méthode 1).

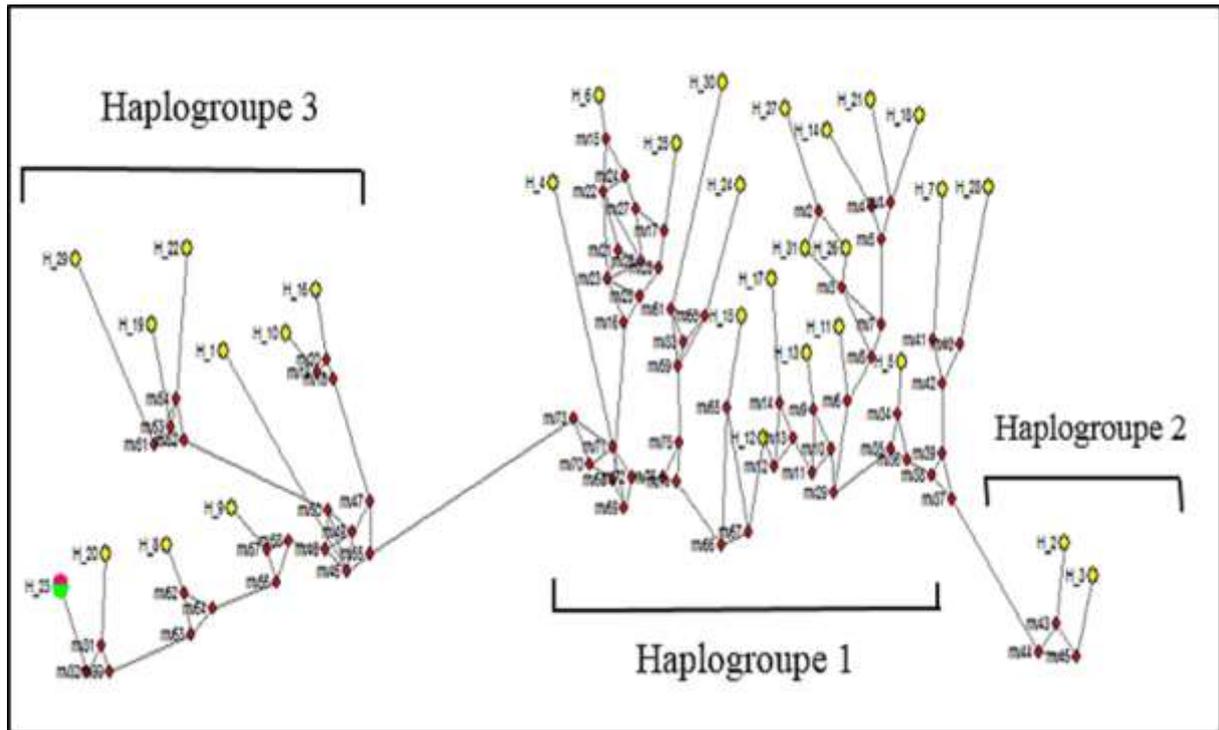


Figure 42: Réseau haplotypique des séquences de gènes RdRp réalisé par le logiciel NetWork basé sur les souches virales de coronavirus dans différentes régions géographiques.

III.1.5.2 Résultat 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 différentes souches de coronavirus qui ont infecté différentes espèces animales à travers plusieurs pays dans le monde

Pour le gène S (**Figure 43**) nous avons obtenu quatre groupes

Alpha CoV comprend les souches suivantes : CCOV de la Chine, FeCoV de la Chine, FCoV de la Chine, PRCV de l'Espagne, TGEV de la Chine, PEDV de la Chine, PEDV du Vietnam qui sont présentés avec couleur unique des cercles (jaune) chaque souche présentée par une seule séquence, à l'inverse dans les cercles avec des couleurs différentes l'inclusion de deux séquences pour les souches PEDV du Canada et PEDV du Japon (cercle mauve et bleu) et PEDV de Thaïlande et du Mexique (cercle vert et gris).

Delta Cov comprend les souches suivantes : PDCoV de la Chine, PDCoV du Mexique, PEDV de l'USA.

Gamma CoV comprend les souches suivantes : TCoV des USA, gfCoV de la France.

Beta CoV comprend les souches suivantes : SARS-CoV2 des Usa, SARS-CoV2 de l'Iran, SARS-CoV1 de la Chine, MERS-CoV de l'Arabie Saoudite, Bat Cov de la Chine, HKU1 de Pays bas, HKU1 de la Chine, RtBs de la Chine, DcCoV du Maroc, DcCov de l'Ethiopie, HCoV OC43 de la France, BCoV de la Corée du Sud, BCoV de la Chine, Yak CoV de la Chine.

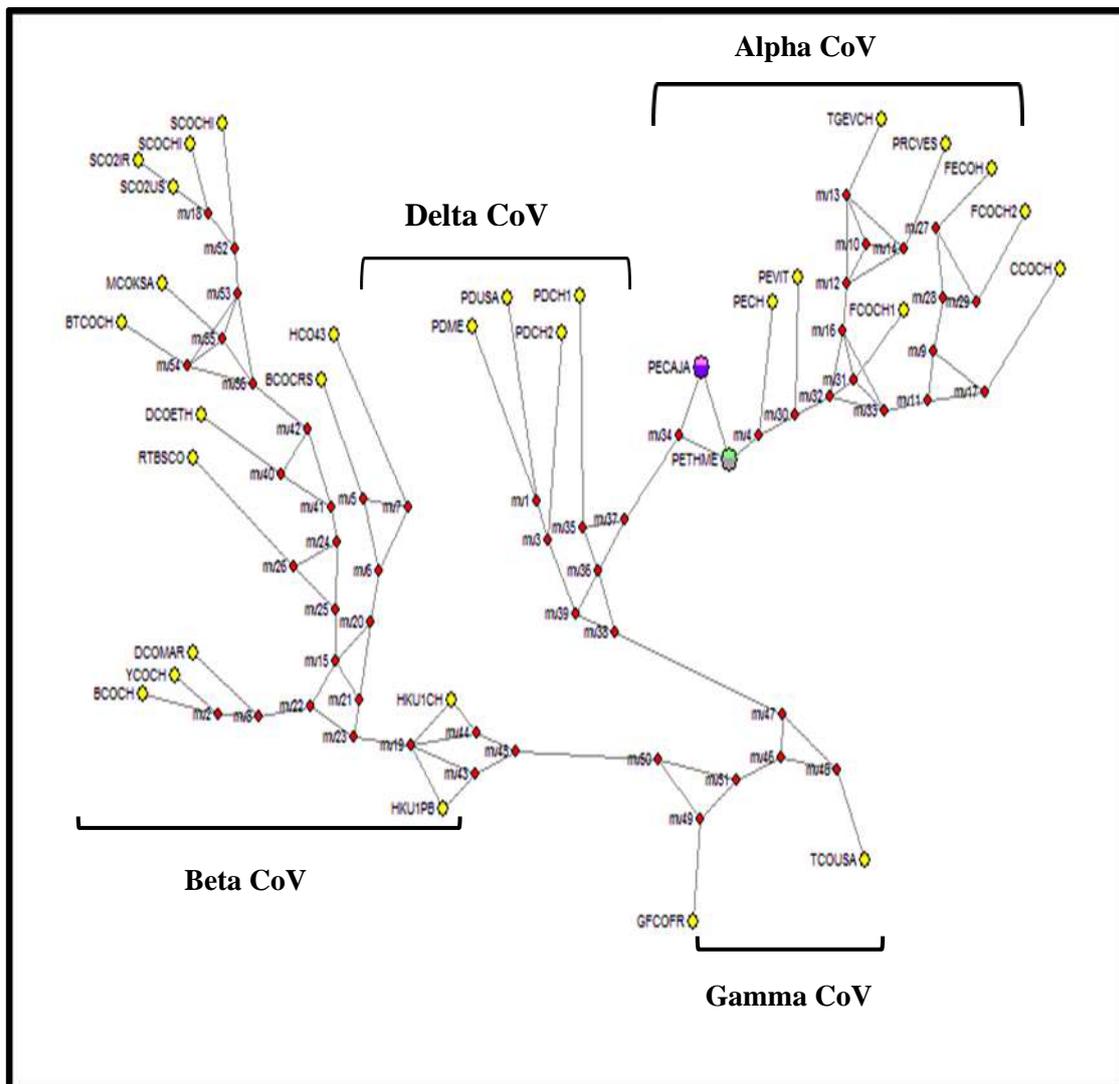


Figure 43 : Réseau haplotypique des séquences de gène S réalisé par le logiciel NetWork basé sur les souches virales de coronavirus dans différentes régions géographiques.

**CO (Coronavirus), CCO (CCoV), FCO (FCoV), FECO (FECoV), PE (PEDV), PD (PDCoV), Dc (DcCoV), S (SARS), SCO2 (SARS-CoV2), M (MERS), BT (Bat), PRCV (Coronavirus respiratoire porcin), TGEV (Coronavirus de la gastro-entérite transmissible porcine), GF (Guinea fowl pintade), B (Bovin), Y (Yak), HKU1 (Coronavirus humain de Hong kong), T (Coronavirus de Turquie infecte le dinde), Rt-Bs (Rat Bandicota savilei), HCO43 (Humain coronavirus OC43), CH (Chine), ES (Espagne), VIT (Vietnam), CA (Canada), JA (Japon), TH (Thaïlande), ME (Mexique), P (Pays-Bas), FR (France), MA (Maroc), IR (Iran), ETH (Ethiopie), KSA (Arabie Saoudite), CRS (Corré du Sud), US (USA).

Pour le gène RdRp (**Figure 44**) nous avons obtenu trois groupes:

Alpha CoV comprend les souches suivantes : BtCov du Brésil, Japon, Madagascar, PEDV de l'Italie, RtCov de la Chine, HCov/229E du Brésil et de la Slovénie.

BetaCoV renferme le BtCov de l'Arabie Saoudite, de la Chine, Mozambique, Kenya, Corée du Sud, Inde, Espagne, le RtCov de Chine et du Vietnam en ce qui concerne les rongeurs, PHEV de l'Italie.

GammaCov renferme les deux ACov de la Pologne.

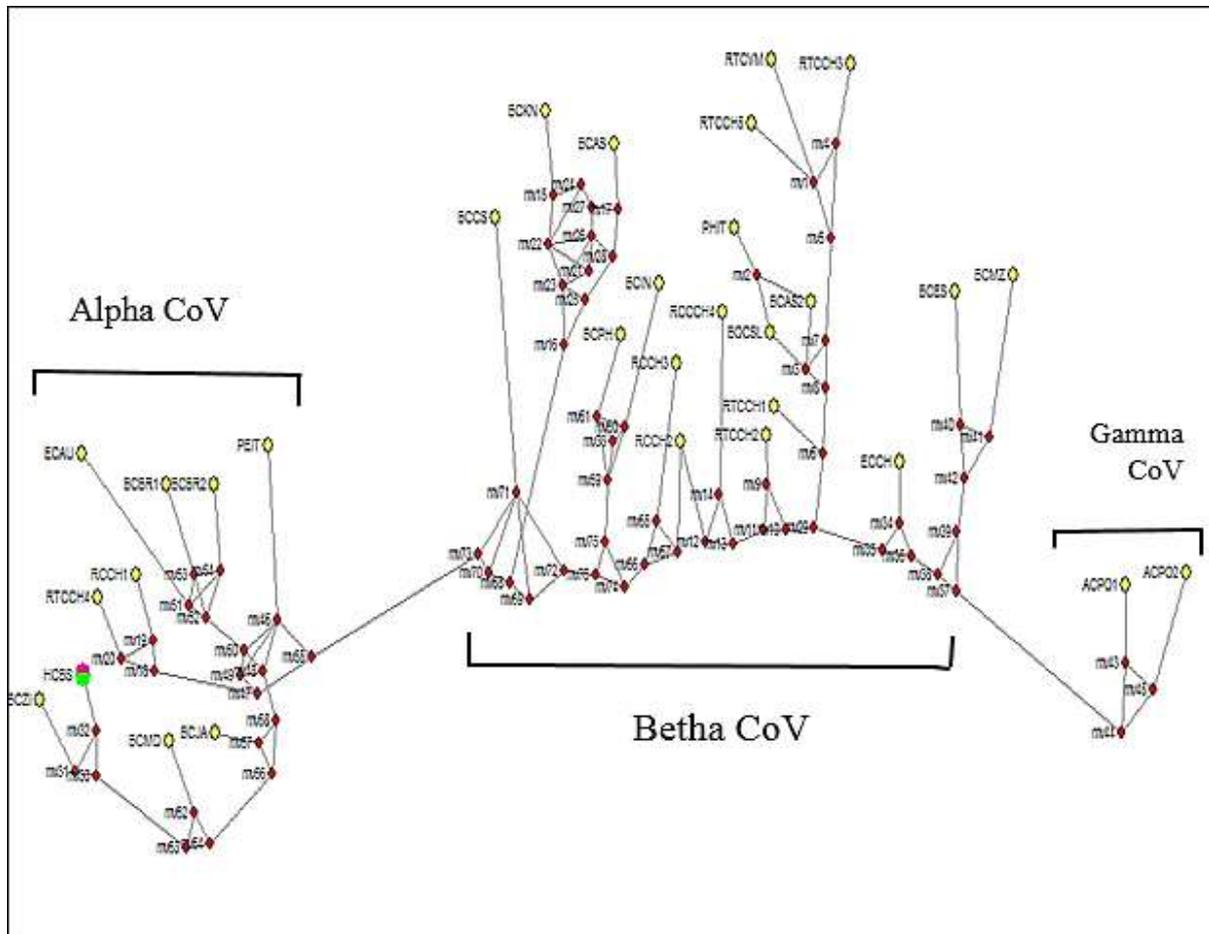


Figure 44 : Réseau haplotypique des séquences de gène RdRp réalisé par NetWork basé sur les souches virales du coronavirus dans différentes régions géographiques (méthode 2).

**PE (PEDV), AC (Avian Cov), BC (BtCov), EC (ECov), RC (RtCov, souris), Rtc (Rtcov, rat), FC (Frscv), BOC (Bcov), BR (Brésil), VM (Vietnam), KN (Kenya), IN (Inde), AU (Australie), MZ (Mozambique), ES (Espagne), IT (Italie), AS (Arabie Saoudite), PH (Philippines), CH (Chine), CS (Corée du Sud), MD (Madagascar), JA (Japon).

III.2. Discussion

III.2.1. La phylogénie du coronavirus

III.2.1.1 Les relations phylogénétiques au sein de l'arbre du gène S

Notre arbre phylogénétique se compose de quatre groupes Alphacoronavirus, Betacoronavirus, Gammacoronavirus et Deltacoronavirus ils contiennent diverses souches virales, et infectent différentes espèces animales. D'après **Woo et al., (2012)** les

alphacoronavirus et bêtacoronavirus n'infectent que les mammifères, les gammacoronavirus et les deltacoronavirus infectent les oiseaux, mais certains d'entre eux peuvent également infecter les mammifères.

Selon les travaux de **Lefkowitz et al., (2011)** le coronavirus de canin CCoV qui infecte le chien est étroitement lié au virus de la gastro-entérite transmissible (TGEV) du porc, au coronavirus du furet et au coronavirus félin (FCoV). Dans notre arbre le CCoV de la Chine qui infecte le chien, PRCV de l'Espagne et TGEV de Chine infectant le porc, FCoV et FeCoV de la Chine qui infecte le chat, sont regroupés dans le même sous-groupe Alpha a (α). La souche CCoV (HCM47/2015) et FECoV (DM95/2003) sont dans la même branche ces résultats sont expliqués par **Dung et al., (2017)** ou ils ont trouvé par l'analyse phylogénétique de gènes S complet que CoV/HCM47/2015 de la Chine était similaire aux CoV/DM95/2003, ils partagent 95,9% de similarité ainsi d'après **Yuan et al., (2018)** le coronavirus respiratoire porcine (PRCoV) est une variante respiratoire du TGEV, la protéine S codée par PRCoV manque d'environ 200 acides aminés dans la région N-terminale qui contiennent des déterminants liés à l'entéropathogénicité du TGEV pour cela la souche TGEV de la Chine et PRCV de l'Espagne sont dans la même branche.

Au sein de notre arbre phylogénétique le PEDV du Japon (14JM-01), le PEDV du Canada (ON-007), PEDV de Thaïlande (P1915-NPF-071511A), PEDV de Mexique (PEDV/MEX/MICH/01/2013), PEDV de la Chine (SHpd/2012), PEDV du Vietnam (HUA-PED88) qui infectent les porcs appartiennent au même sous-groupe (α). **Nguyen et al., (2018)** ont montré que la souche du Japon 14JM-01 est une souche de type nord-américain à 100% d'identité dans la séquence nucléotidique de gène S avec la souche canadienne (ON-007), ainsi ils citent que la souche P1915-NPF-071511A est la première souche de l'émergence d'un PEDV de type américain en Thaïlande, la séquence de gène S partage 100% de similarité avec les souches américaines KUIDL-PED-2014-002, USA/Ohio60/2013 qui partagent avec la souche du Mexique (PEDV/MEX/MICH/01/2013) 99.81% de similarité, ces travaux confirment notre résultat pour les quatre souches précédentes qui étaient dans la même branche, en parallèle **Than et al., (2020)** montrent que la souche HUA-PED88 du Vietnam/2015 est génétiquement liée aux souches PEDV circulant en Chine SHpd/2012.

D'après les travaux de **Qingqing et al., (2019)** la transmission interspécifique potentielle du PDCoV peut exister entre les porcs domestiques et les oiseaux/aviaires, indiquant que le coronavirus PDCoV CH/GD03/2015 de porc a infecté les poulets (PDCoV CC-HNZK) et la PDCoV CC-HNZK-02 présentait 2 mutations dans le gène S : Asp est devenu His à la position 138 et Gln a été remplacé par Lys à la position 641 ce qui explique la position de

PDCoV CC-HNZK proche de PDCoV CH/GD03/2015 dans notre arbre, ainsi **Pérez et al., (2019)** ont cité que la souche mexicaine PEDV est phylogénétiquement plus proche des souches PEDV signalé aux États-Unis, l'analyse des séquences du gène S obtenu à partir d'isolats mexicains partageaient 99,6% d'identité nucléotidique par rapport aux souches nord-américaines.

Le coronavirus de pintade GfCoV/FR/2011 de la France et le coronavirus de l'Inde TCoV (IN/421/10) et de l'USA sont dans la même branche, selon **Kim et al., (2019)** le coronavirus de pintade GfCoV/FR/2011 de la France est étroitement associé au coronavirus de l'Inde TCoV (IN/421/10) avec une similarité de 86.44%.

Le SARS-CoV-2 signalé pour la première fois en Chine en décembre 2019 (**Zhou et al., 2020**), est rapidement devenue une pandémie aux effets dévastateurs. Selon **Manuel et al., (2020)** la variante portant la mutation D614G Spike qui est la transition d'une aspartate (D) à la position 614 trouvé fréquemment dans les souches chinoises à une glycine (G). D'après **Korber et al., (2020)** les voyageurs ont dispersé les variantes du G614 dans le monde et auraient probablement introduit et réintroduit des variantes du G614 dans différents endroits, l'épidémie asiatique était majoritairement D614 jusqu'en février. Début avril 2020, G614 est depuis devenu la forme dominante de la pandémie ce qui explique la position des souches SARS-CoV-2/human/IRN/K1r-112/2020" de l'Iran et SARS-CoV-2/human/USA/TG701123/2020 des Etats Uni dans le même branchement, les 2 collecté après avril 2020 contiennent la mutation du G614.

Les deux souches SARSr-CoV de civette de la Chine sont dans la même branche et partage une similarité de 99.42%, selon **Guan et al., (2003)** après l'émergence de coronavirus humaine syndrome respiratoire aigu sévère (SRAS-CoV) dans le sud de la Chine en 2002/2003, ont été trouvé le type de SRAS-Like (SARSr-CoV) porté par des civettes de palmiers provenant des marchés d'animaux vivants du Guangdong, leurs séquences génomiques présentaient une identité de 99,8% avec celle du SARS-CoV humain, elles étaient considéré comme hôte intermédiaire qui a permis le passage du virus à l'homme .

Selon **Guangwen et al., (2015)** le MERS-CoV appartient à la lignée C des bêtacoronavirus, étant étroitement liés à Tylonycteris bat CoV HKU4 (Ty-BatCoV HKU4) dans le gène S (66,8% à 67,4%), des fragments de gènes viraux identiques ou assez similaires à ceux du MERS-CoV ont également été récupéré chez les chauves-souris, soulevant à nouveau la possibilité que la chauve-souris agisse comme le réservoir naturel du MERS-CoV, ainsi **Chen**

et *al.*, (2013) déclarent que la majeure partie de la recombinaison et de la diversité des séquences se trouve dans le gène S ces changements peuvent affecter la conformation structurelle du RBD (les sites de sélection positif se trouvaient dans le RBD de la protéine S, RBD montre deux parties, une région de liaison et une région centrale. Les deux sont cruciales pour que le virus reconnaisse et pénètre dans les cellules hôtes et interagisse avec les récepteurs humains apparentés. Le MERS-CoV a utilisé le récepteur DPP4 pour pénétrer dans les cellules hôte. Il a été observé sur deux sites de mutations dans la région de liaison de RDB, suggérant que ces substitutions d'acides aminés pourraient améliorer la capacité du MERS-CoV à se lier à différentes cellules hôtes et ainsi faciliter sa transmission entre espèces (Zhang et al., 2016). Les fortes similitudes entre les DPP4 des humains, des chameaux, des chevaux et des chauves-souris permettent au MERS-CoV de les infecter (Folegatti et al., 2021). Cela justifie la position des souches MERS CoV de l'Arabie Saoudite et Tylonycteris bat CoV HKU4 de la Chine dans la même branche.

Ray et al., (2019) ont cité que le DcCoV-HKU23 du Maroc qui a touché le chameau diffère du BCoV infecte le Bovin par une plage de 0,0249 à 0,0300 substitutions de bases de paires par site et de HCoV-OC43 de l'humain par une plage de 0,0445 à 0,0468 substitutions de bases de paires par site, également l'espèce la plus proche du HCoV-OC43 reste le BCoV plutôt que le DcCoV-HKU23, BCoV et HCoV-OC43 partagent une identité nucléotidique globale de 96% dans leur gène S, ce qui justifie la position de la souche BCoV de la Corée du Sud qui est la plus proche de la souche de l'espèce humaine OC43 (HCoV-OC43) de la France dans notre arbre ; ainsi selon les travaux de Nathalie et al., (2016) sur la comparaison génétique du BCoV et du HCoV-OC43 révèle que le HCoV-OC43 résulte d'une transmission zoonotique du bovin à l'homme. Etant donné que ces deux virus sont extrêmement proches en ce qui concerne les caractéristiques génétiques et épidémiologiques.

III.2.1.1.2 Les relations phylogénétiques au sein de l'arbre du gène Rdrp

Selon les travaux de Nischay Mishra et al. (2019) sur les données métagénomiques virales des chauves-souris dans l'Arabie Saoudite (KSA), la séquence de KSA302, récupérée à partir d'un *T. perforatus* bat, avait 100% d'identité avec un coronavirus bovin de Louisiane, trois produits de 400 nucléotides supplémentaires de *E. helvum* les chauves-souris avaient une homologie de 99 à 100% nucléotides avec un bêtacoronavirus de chauve-souris Roussetus HKU9 de la Chine ce qui confirme nos résultats concernant le classement de *T. perforatus* KSA302 (betacoronavirus). Cependant en 2012 le Royaume de l'Arabie Saoudite a signalé un

cas de pneumonie mortelle associée à un nouveau coronavirus, une infection par le MERS-CoV. Alors que les principaux cas humains de MERS-CoV sont souvent associés à une exposition épidémiologique à des dromadaires infectés, les analyses phylogénétiques et la distribution mondiale des virus liés au MERS-CoV chez les chauves-souris suggèrent qu'ils étaient probablement une source originale de propagation aux chameaux.

Concernant la position de la souche de coronavirus BCoV qui infecte le bovin en Slovénie dans notre arbre phylogénétique, selon **Monika Jevsnik et al., 2021** le BCoV est une souche provoque une diarrhée néonatale de veau et une dysenterie hivernale chez les animaux adultes, est une souche transmissible entre espèce et il a été détecté chez plusieurs espèces de ruminants et chez l'homme, toutes les souches caractérisées de différentes zones géographiques appartiennent au sous-groupe initialement désigné comme 2a, et selon la nouvelle classification de ICTV, BCoV appartient au genre Betacoronavirus, donc malgré le changement géographique de la souche BCoV reste toujours classé dans ce genre, sa position dans l'arbre phylogénétique du gène S confirme sa classification.

Leopardi et al., (2020) ont publié que un et sept échantillons étaient positifs chez les chauves-souris et les porcs, respectivement. Toutes les souches ont été classé sur la base de séquences partielles de RdRp et ressemblaient fortement à des espèces connu déjà décrites dans la même espèce hôte, en détail tous les Cov trouvés chez les porcs étaient associés à des Cov porcins connu, à savoir le virus de l'encéphalomyélite hémagglutinante porcine PHEV et le virus de la diarrhée épidémique porcine PEDV. Cependant il a été confirmé que les chauves-souris volent fréquemment au-dessus des élevages porcins intensifs italiens avec une biodiversité et une activité élevée. En effet, on a pu enregistrer 15 espèces, représentant plus de la moitié trouvée dans le nord-est de l'Italie. Ces données sont cohérentes avec les découvertes d'habitats précédemment considérés comme beaucoup plus attrayants pour les chauves-souris, tels que les bois ou le maquis méditerranéen. Il est probable qu'un résultat aussi inattendu soit lié à l'agrégation d'insectes autour des réservoirs d'eaux usées, une hypothèse qui est étayée par l'activité fourragère intense jusqu'à 72% enregistré. Aussi ils ont systématiquement trouvé un regroupement phylogénétique des virus porcins (ou au moins des séquences consensus) sur le lieu de l'échantillonnage, ce qui signifie que la plupart des événements de transmission se produisent au sein d'une seule ferme. Cependant, la corrélation constante des virus des élevages 2 et 4 met en évidence les risques de dispersion des agents pathogènes par les mouvements d'animaux dans les cycles ouverts de production porcine. En comparant ces données avec nos résultats concernant l'arbre et le réseau et le pourcentage de

similitude proches entre la souche virale qui touche les ch-s qui est BtCov du Japon et de Madagascar, et la souche PEDV qui touche le porc en Italie ce que signifie leur position ensemble dans un même clade et même haplogroupe. En plus les élevages des porcs qui ont un contact avec les chauves-souris et la grande capacité de vol de ces derniers augmente la probabilité d'infection chez l'humain par le virus donc ce qui suggèrent selon notre arbre et le réseau la position proche de la souche virale HCov/229E, qui infecte l'homme au Brésil et en Slovénie et la souche virale BtCov infectant le ch-s au Zimbabwe et Madagascar.

Selon les résultats obtenu dans l'arbre phylogénétique et le réseau d'haplogroupe, la souche de coronavirus Rt-CoV qui touche les rongeurs trouvée dans les pays d'Asie, et selon les travaux de **Zhiqiang Wu et al., 2018** sur le génome des rongeurs, ces animaux qui représentent environ 43% de toutes les espèces de mammifères ; ils sont largement répartis, ils représentent les réservoirs naturels d'un groupe diversifié de virus zoonotiques, ce qui explique leur présence considérable dans le clade alphacoronavirus et bethacoronavirus.

D'après les recherches de **Lescano et al., 2015**, qui démontre que des furets sont importés comme des animaux de compagnie depuis fin 2012, la population actuelle serait de moins de 250 individus. Le furet impliqué dans ce cas est né en Espagne et a eu une période de quarantaine cumulée (au Chili et au Pérou) d'environ 60 jours. Par la suite, il a vécu dans la ville de Lima pendant environ 90 jours supplémentaires sans montrer aucun signe clinique. Il n'y aurait eu aucun contact avec d'autres furets que le groupe importé. Cela suggère que l'infection s'est produite en Espagne et que l'animal est resté cliniquement asymptomatique pendant environ 150 jours. **Garner et al., 2008 ; Martínez et al., 2008 ; Johnson-Delaney, 2010; Murray et al., 2010** ont déclaré qu'il a été infecté au FRSCV c'est la souche de coronavirus diagnostiqué chez un furet importé d'Espagne au Pérou. C'est la première fois qu'une infection systémique à coronavirus chez les furets est décrite en Amérique latine. Le furet avait des valeurs hématologiques et des signes cliniques similaires à ceux décrit pour la maladie associée au FRSCV. Cependant, ces signes sont relativement non spécifiques et de nombreux autres signes cliniques associés à la FRSCV n'ont pas été observé, notamment des masses intra-abdominales palpables, une splénomégalie, une néphromégalie, un bruxisme, des éternuements/écoulement nasal, un souffle cardiaque systolique, une décoloration des urines, une dyspnée, lymphadénomégalie, érythème de la muqueuse rectale, convulsions, opisthotonus, vomissements et pyrexie. En comparant avec notre résultat en ce qui concerne la position de FRSCV (une seule) dans le clade de alphacoronavirus il n'y avait aucune autre souche présente en Espagne dans ce clade.

III.2.2. Comparaison générale entre les logiciels utilisés

Après l'analyse des gènes de différentes souches de coronavirus par les logiciels : MEGA 5, DnaSP 6, DAMBE et NetWork 10, nous avons remarqué que les deux gènes étudiés S, RdRp représentés par des arbres phylogénétiques donnés par la méthode UPGM (**Figure 33, 35**) sont des gènes avec un taux de variation élevé. **Pedro et al., (2021)** ont montré une accumulation des mutations dans la séquence et la structure de la glycoprotéine de pointe (S) des coronavirus. En parallèle les réseaux phylogénétiques de ces gènes construits par logiciel NetWork représentés dans les figures (**41, 42, 43, 44**) dans lesquels on voit des cercles soit avec couleur unique 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), 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). **Anderson et Karel, (2011)** ont montré que la distance est exprimée par la présence des mutations et traduite par la détection de plusieurs haplotypes pour le même gène ce qui est défini par une diversité haplotypique (H), que nous avons détecté grâce au logiciel DnaSp, nous avons trouvé la diversité haplotypique (Hd) de gène S et RdRp (0.9962 et 0.9983).

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) indiquent que les 2 gènes ont une variation élevée, nous avons remarqué que GC global skew calculé de la plupart des séquences de gènes étudiés est très faible (**Tableau IV**) par rapport à global AT skew (les séquences riches en cytosine par rapport à la guanine et en thymine par rapport à l'adénine), selon **Qin et al. (2015)** une teneur plus faible en C+G indique généralement que la séquence nucléotidique du virus est instable ou très variable.

Conclusion

Le coronavirus est l'un des virus qui fait partie de la famille de coronaviridae, c'est le responsable de syndrome respiratoire aigu chez certains espèces et le syndrome de diarrhée aiguë chez d'autres espèces ce virus a différentes souche tel que : BtCov, RtCov, BCov, ACov, ECov, FECov, PEDV, CCov, HCov/229E, SARS-Cov, MERS-Cov, SARS-Cov2.. Avant 2003 il y'avait que 10 coronavirus avec des génomes complets disponibles, après l'épidémie de SRAS, jusqu'en décembre 2008 il y'a eu un ajout de 16 coronavirus avec des génomes complets séquencés. Ils comprennent les deux coronavirus humain NL63 et HKU1, et les dix coronavirus de mammifères (le SARS-Cov des chauve-souris et le BtCov de chauve-souris aussi, HKU2, bat-CoV HKU4, bat-CoV HKU5, bat-CoV HKU8, bat-CoV HKU9, bat-CoV 512/2005, bat-CoV, le coronavirus d'équin, les 4 coronavirus des aviaires. Arrivant à l'intérêt de notre étude qui a été fondé sur la classification et les liens entre les souches de coronavirus dans différentes zones géographiques nous avons utilisé la phylogénie qui est une méthode utile aussi dans la génétique des populations et l'épidémiologie. C'est une composante essentielle dans la biologie pour l'étude des liens et l'histoire évolutive des organismes et la diversification sur le plan moléculaire et morphologique.

Pour la réalisation de ce travail, nous avons utilisé des outils bioinformatiques permettant l'analyse phylogénétique (la construction de l'arbre phylogénétique) et déterminer les haplotypes, pour arriver à savoir l'évolution de ce virus. Les quatre logiciels sont utilisés en ordre comme suit : MEGA5 pour la réalisation de l'arbre phylogénétique qui comporte les souches virales, DnaSP pour l'identification des haplotypes des souches, le NetWork pour savoir l'évolution de virus, DAMBE pour les calculs statistiques. Ces analyses sont appliquées sur 35 séquences de gène Rdrp et 33 séquences de gène S pour classer les souches virales dans chaque genre selon l'espèce touchée et la zone géographique où ils ont été trouvés.

Les résultats obtenus par le DnaSP sur les séquences de gène ARN-polymérase ARN-dépendant et le gène de protéine de pointe(S) sont compatibles et confirment dans une large mesure les résultats de logiciel MEGA5. Les séquences identiques de chaque haplotype de différentes souches correspondent au même clade dans l'arbre phylogénétique. Selon les résultats de MEGA5 ainsi que NetWork les séquences d'ADN du gène ARN-polymérase-ARN-dépendant (Rdrp) étudié se regroupent en 3 haplogroupes représentés par trois clades différents et quatre haplogroupes représentés aussi par quatre clades différents pour le gène de protéine de

pointe(S), les espèces animales touchés se diffèrent ainsi que les souches virales selon le changement géographique .la large répartition de chauve-souris dans le monde et la représentation de 43% de rongeurs des mammifères surtout en Asie justifie la présence de leur souches virales dans les deux clades alpha et bêta, la présence de 2 séquences de Acov dans le clades de Gamma coronavirus ,et les autres souches classées auparavant ceci montre qu'il y'a une grande diversité génétique entre les souches de coronavirus qui est due à l'infidélité de gène RdRp et S c'est-à-dire leurs taux de mutations est très élevé.

Références bibliographiques

- **Alauzet C., 2009.** Taxinomie des bactéries anaérobies : de la reclassification à la découverte de nouveaux pathogènes. Thèse doctorat, université de Nancy.
- **Alluwaimi A. M., Alshubaith I. H., Abohlaïka S., 2020.** Les coronavirus des animaux et des oiseaux : leur zoonose, leurs vaccins et leurs modèles pour le SARS-CoV et le SARS-CoV2, 7: 582287.
- **Ana F., Sonia V-M, Maisons I., Caroline A., Guillermo R., Francisco P., Pilar P-B., Javier Juste , Carlos I., Inazio G., Joxerra A , Juan E ., 2011.** Détection des alpha et bêta coronavirus chez plusieurs espèces de chauves-souris ibériques. Archives de virologie 156(10): 1883-90.
- **Biniotti M., Paptti A., Lavazza A., Alborali G., Sozzi E., Chiapponi C., Facini S., Bonilauri P., Codioli P., Marthelar D., 2016.** Virus de la diarrhée épidémique porcine et découverte d'un coronavirus entérique porcine recombinant, Italie, Centers for Disease Control and Prevention, 22(1) :83-7.
- **Blanchet C., 1999.** Logiciel MPSA et ressources bioinformatique client -serveur Web dédiés à l'analyse de séquences de protéine. Thèse de Doctorat en Bioinformatique. Dir Thèse : Delage G. Univ. Claude Bernard – Lyon 1 :23-36.158 p.
- **Bourgarel M., Pfeukenyi D., Boui V., Talignani L., Chiwechi N., Diop F., Caron A., Matop C., Missé D., Léogois F., 2018.** Circulation des Alphacoronavirus, Betacoronavirus et Paramyxovirus chez les espèces de chauves-souris *Hipposideros* au Zimbabwe. Revue d'épidémiologie moléculaire et de génétique évolutive des maladies infectieuses ,58 :253-257.
- **Bouwman K.M., Delpont M., Broszeit F., Berger R, Weerts E.A., Lucas M.N., Delverdier M., Belkasmi S., Papanikolaou A., Boons G.J., Guérin CL., Robert P., Ducatez M.F., Verheije M.H., 2019.** A diversité des coronavirus de pintade a des conséquences phénotypiques pour la liaison au glycane et aux tissus. Journal of virology e00067-19.
- **Chen N., Zhou M., Dong X., Qu J., Gong F., Han H., Qiu Y., Wang J., Liu Y., Wei Y., Xia J., Yu T., Zhang X., Zhang L., 2020.** Caractéristiques épidémiologiques et cliniques de 99 cas de pneumonie à nouveau coronavirus 2019 à Wuhan, Chine. Lancette 395 (10223) : 507-513.

- **Chen Y.N., Loa C.C., Ababneh M.M.K., Wu C.C., Lin T.L., 2015.** Génomique d'isolats de champ de coronavirus de dinde provenant de divers emplacements géographiques aux États-Unis sur la base du gène de pointe, 160(11) : 2719-2726.
- **Chiekh M., 2006.** Évaluation des classifications phylogénétiques des Bacillaceae basées sur les gènes de l'opéron *rrn* et de gènes de ménage. Université de Québec à Montréal.
- **Chiu S., Chan K.H., Chu K.W., Kwan W., Guan Y, Poon L.L.M., Peiris J.S.M., 2005.** Infection au coronavirus humain NL63 et autres infections à coronavirus chez des enfants hospitalisés pour une maladie respiratoire aiguë à Hong Kong, Chine. *Clinical infectious diseases* 40(12) : 1721-1729.
- **Choi J. H., Jung H. Y., Kim H. S., Cho H. G., 2000.** PHYLODRAW: A Phylogenetic Tree Drawing System. Department of computer science, Pusan National University, Pusan, Korea.
- **Clark k., Mizrachi.I.K., Lipman D.J., Ostell J., Sayers E.W., 2016.** Genbank Nucleic Acids Research D67–D72.
- **Corman V. M., Muth D., Niemeyer D., Drosten C., 2018.** Hôtes et sources de coronavirus humains endémiques, 100 : 163-188.
- **Cui J., Li F., Shi Z.L., 2019.** Origine et évolution des coronavirus pathogènes. *Nat Rev Microbiol*, 17(3) : 181–192.
- **Darlu P., Tassy P., 2008.** La reconstruction phylogénétique : concepts et méthodes. Ed Masson, Paris.P1 et 2 et 3.
- **Dasilva L., Zerbinati R., Tateno A., Boas L., Almeda M., Levi J., Dexelar F., Drosten C., Pannuti C., 2012.** L'impact clinique différentiel des espèces de coronavirus humain chez les enfants atteints de mucoviscidose, *Le Journal des maladies infectieuses*, 206(3) : 384-8.
- **Diallo A. B., 2009.** Bioinformatique avancée. Université du Québec à Montréal. Canada.
- **Deng X., Baker S., 2021.** Coronavirus : biologie moléculaire (Coronaviridae). *Encyclopédie de virologie* 198-207.
- **Dung V.N., Terada Y., Minami S., Yonemitsu K., Nagata N., Thanh L.E., Kuwata R., Shimoda H., Maeda K., 2017.** Caractérisation de la propagation du coronavirus canin chez les chiens domestiques au Vietnam. *The Journal of veterinary medical science* 79(2): 343-349.

- **Felsenstein J., 2004.** Inferring phylogenies. Sunderland (MA): Sinauer Associates: 664.
- **Folegatti P.M., Bittaye M., Flaxman A., Dphil, Ramos L.F., Bellamy D., Kupke A., Mair C., Makinson R., Sheridan J., Rohde C., Halwe S., Jeong Y., Park Y.C., Kim J.K., Song M., Boyd A.T. N., Silman D., Poulton I., Dattoo M., MRCP., Marshall J., Themistocleous ., Lawrie A., Roberts R., Berrie E., Becker S., Lambe T., Hill A., Ewer K., Gilbert S., 2020.** Innocuité et immunogénicité d'un vaccin à vecteur viral contre le coronavirus du syndrome respiratoire du Moyen-Orient : un essai de phase 1 en ouvert, non randomisé, non contrôlé. *La lancette maladies infectieuses* 20(7) : 816-826.
- **Fournel M., Tardif C., 1834.** Traitee du voisinage considaerae dans l'ordre judiciare et dans ses rapports avec le code civil. Volume 1.Ed. Videcoq, Paris .52p.
- **Fung TS., Liu DX., 2019.** Coronavirus humain : interaction hôte-pathogène. *Annual review of Microbiology*, 73:529-557.
- **Gao M., Chang C., Xie Y., Tain G., Jian G., Jiang H., Téo C., Shen J., Xu Y., Zheng S., Zhang Y., Xu H., 2020.** Base structurelle de l'inhibition de l'ARN polymérase ARN-dépendante du SARS-CoV-2 par le remdesivir,368(6498) :1499-1504.
- **Gattolliat J. L., 2002.** Etude systématique, cladistique et biogéographique des Baetidae (Ephemeroptera) de Madagascar, Thèse de doctorat en Zoologie et Ecologie.
- **Golding B., Morton D., 2003.** Elementary sequence Analysis. Department of Biology. Mc Master.Univ. Hamilton. Ontario .190p.
- **Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M., 1989.** Génome du coronavirus : prédiction de domaines fonctionnels putatifs dans la polyprotéine non structurelle par analyse comparative de séquences d'acides aminés. *Acides nucléiques Resaerch* 17(12) : 4847–4861.
- **Guan Y., Zheng BJ., Liu XL., Zhuang ZX., Cheung CL., Luo SW., LJ Zhang., Fesses KM., Wong KL., Chan KW., Lim W., Shortridge., Yuen KY., Peiris JSM., 2003.** Isolement et caractérisation des virus liés au coronavirus du SRAS chez les animaux du sud de la Chine.*Jornal of science* 302(5643) :276.
- **Guan Y., Zheng Y., Zhuang C., Cheung S., Luo W., Li H., Zhangyj G., Wongkw F., shortridgeky L., Yuenjism P., 2003.** Isolement et caractérisation des virus liés au

coronavirus du SRAS chez les animaux du sud de la Chine. *Journal of science* 5643(276 – 27) p.

- **Guangwen L., Hu Y., Wang Q., Qi J., Gao F., Li Y., Zhang Y., Zhang W., Yuan Y., Bao J., Zhang B., Shi Y., Yan J., Gao G.F., 2003.** Base moléculaire de la liaison entre le nouveau coronavirus humain MERS-CoV et son récepteur CD26, *500(7461): 227-231.*
- **Hendley J., Fishburne H.B., Gwaltney J.M., 1972.** Infections à coronavirus chez les adultes qui travaillent. Étude de huit ans avec 229 E et OC 43, 805-11.
- **Jaimes J.A., Whittake J. R., 2018.** Coronavirus félin : aperçu de la pathogenèse virale basée sur la structure et la fonction de la protéine de pointe. *Virologie 517 : 108-121.*
- **Kin N., Miszczak F., Diancourt L., Caro V., Moutou F., Vabret A., Gouilh M.R., 2016.** L'épidémiologie moléculaire comparative de deux coronavirus étroitement liés, le coronavirus bovin (BCoV) et le coronavirus humain OC43 (HCoV-OC43), révèle un modèle évolutif différent, 186-191.
- **Korber B., Fischer W.M., Gnanakaran S., Yoon H., Theiler J., Abfalterer W., Hengartner N., Giorgi E., Bhattacharya T., Foley B., Hastie K.M., Parker M.D., Partridge D.J., Evans C.M., Freeman T.M., Thushan I., Sheffield COVID-19 Genomics Group, Danal C., 7 Perez L.G., Tang H., Walker A.M., Whelan S., Celia C., Saphire E., Montefiori D.C., 2020.** Suivi des changements dans le pic de SARS-CoV-2 : Preuve que le D614G augmente l'infectiosité du virus COVID-19. *Revue Cell 182(4) : 812-827.e19.*
- **Kuhls k.,Mauricio I., 2019.** Phylogenetic studies. *Molekulare Biotechnologie und Funktionelle Genomik, 1971:968.*
- **Lehmann K.C., Hooghiemstra L., Gouliaeva A., Samborskiy D.V., Zevenhoven J.C., Snijder E.J., Gorbalenya E.J., Clara C., Posthume C.C., 2015.** Artérovirus nsp12 versus coronavirus nsp16 2'-O-méthyltransférase : comparaison des produits de clivage C-terminal de deux polyprotéines de nidovirus pp1ab. *Journal de virologie générale 96 (9) : 2643-2655*
- **Lemey P., Salemi M., Vandamme A. M., 2009.** The phylogenetic handbook. Cambridge: Cambridge University Press. Nouveau coronavirus aviaire et maladie fulminante chez la pintade. *Revue Emerg Infect Diseases. 20(1) : 105-108.*

- **Lescano J., Quevedo ., Gonzales Vi., Quille MK., Goergi F., 2015.** Premier cas d'infection systémique à coronavirus chez un furet domestique (*Mustela putorius furo*) au Pérou,62(6) :581-5.
- **Liais E., Croville G., Mariette J., Delverdier M., Lucas M N., Klopp C., Lluch J., Donnadieu C., Guy J S., Corrand L., Mariette F., Ducatez., Guérin J L., 2014**
- **Lu R., Zhao X., Li J., Niu P., Yang B., Wu H., 2020.** Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 566–574.
- **Manuel B.F., Cardozo T., 2020.** La mutation du pic viral G614 du SRAS-CoV-2 présente un taux de létalité plus élevé. *Infections diseases* 10.1111.
- **Mao N., Müller M.A., Wentao L., Wang C., Corine H. Geurtsvan., Victor M. Corman V.M., Lamers M.L., Sikkema R.S., Bruin E., Chandler F., Yazdan, Descamps D., Fidouh N, Reusken C., Bosch B.J., Drosten C., Marion PG KoBart L., 2020.** Syndrome respiratoire aigu sévère Réponses aux anticorps spécifiques au coronavirus 2 chez les patients atteints de maladie à coronavirus. *Emerging infections diseases* 26(7): 1478-1488.
- **Mishra N., Fagbo S., Aglagaili A., Nitido A., Williams S., Nj J., Lee B., Durosinlorun A., Jain K., Garcia J., Kpoor V., Olival K., Lipkin W., 2019.,** Une étude métagénomique virale identifie des virus de mammifères connus et nouveaux chez les chauves-souris d'Arabie saoudite, 14(4) : e0214227.
- **Nguyen V.D., Sueyoshi M., Norimine J., Hirai T., Myint O., Teh A.A.P., Izzati Z., Fuke N., Yamaguchi R., 2018.** Caractérisation moléculaire des souches de type américain et asiatique non S INDEL du virus de la diarrhée épidémique porcine (PEDV) qui ont circulé au Japon entre 2013 et 2016 et des PEDV collectés lors d'épidémies récurrentes. *BMC veterinary search* 14: 96.
- **Nijman I. J., Osten M. Verkaar E.I., Rujiter C., Hanekamp E. 2003.** Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) related by mitochondrial DNA, Satellite DNA, AFLP and microsatellites. *Herdity*,90: 10-16
- **Organisation mondiale de la santé, 2021.** Covid-19 en Algérie.
- **Pagel M., 2002.** Discussion comments on Bayesian phylogénétique inference from animal mitochondrial genome arrangement. *Journal of the Royal Statistique Society, serie B, vol.64, Part 4, pp.681-693.*

- **Patric C., Susanna L., Yi H., Kyok Y., 2009.** Diversité des coronavirus phylogénie et saut interspécifique, 234(10) :1117-27.
- **Pérez C.R., Mendoza H.R., Elvira R.S 3, Velázquez R.S., Betancourt J.I., 2019.** Premier rapport et analyse phylogénétique du deltacoronavirus porcin au Mexique.Revue of medecin veterinary 66(4) :1436-1441.
- **Perlman S., Netland J., 2009.** Coronavirus post-SRAS : mise à jour sur la réplication et la pathogénèse. Revues de nature. Microbiologie 7(6) : 439-450.
- **Perrière G., 2000.** Bases de données et outils d'analyse pour la génomique bactérienne. Mémoire de l'habilitation a dériquer des recherches. Univ. Claude Bernard, LYON 1.100p.
- **Philip V., Kratzel A., Steiner S., Stalder H., Thiel V., 2020.** Biologie et réplication du coronavirus : implications pour le SRAS-CoV-2. Clinical infectious diseases 1-16.
- **Pillet V., 2002.** Méthodologie d'extraction automatique d'information à partir de la littérature scientifique en vue d'alimenter un nouveau système d'information et de la communication. Dir. Thèse : L Quoniam et B. Jac.Univ.A.x Marseille 3. 20-43.143 p.
- **Pusterla N., Vin R., Leutenegger C.M., Mittel L.D., 2018.** Divers Infection à coronavirus entérique chez les chevaux adultes, 231 : 13-18.
- **Qingqing Li., Zhang H., Li B., Qingwen D., Wang Y., Gao W., Guo D., Wei Z., Hu H., 2019.** Susceptibilité des poulets à l'infection par le deltacoronavirus porcin. Revue viruses 11(6) : 573.
- **Qin T., Song Y., Shi M., Cheng Y., Zhang W., Xia X., 2015.** Déduire les hôtes du coronavirus à l'aide de modèles statistiques doubles basés sur la composition en nucléotides. Scientifcs reports 5 : 17155.
- **Ranwez V., Gascuel O., 2013.** Méthodes efficaces pour reconstruire de grandes phylogénies suivant le principe de Maximum de vraisemblance. Thèse de doctorat en bioinformatique. Université Montpellier 2- science et Technique du Languedoc 2002.Français : 49-50 .122
- **Ranwez V., 2013.** Méthodes efficaces pour reconstruire de grande phylogénies suivant le principe de maximum de vraisemblance.Thèse de doctorat en bioinformatique .Université Montpellier 2. Sciences et thechniques du Languedoc 2002.Français :49-50.122p
- **Rasmont R., 1997.** Evolution Biologique. Traduction de la 2ème édition anglaise. Ed. Départ. Deboeck Univ. Paris. Bruxelles. 371-507.

- **Ray T. A., Chu D., Miguel E., Perera R., Oladipo J., Fihri O.F., Aylet G., Ko R., Zhou Z., Cheng M.S., Kuranga S.A., Roger F.L., Chevalier V., Webby R.J., Woo D., Poon L.M., Peiris.M., 2019.** La diversité du coronavirus HKU23 du chameau dromadaire chez les chameaux africains a révélé de multiples événements de recombinaison parmi des bêtacoronavirus étroitement liés du sous-genre Embecovirus. *Journal virology* 93(23) : e01236-19.
- **Salemi M., Vandamme A. M., 2003.** The phylogenetic handbook: a practical approach to DNA and protein phylogeny. Cambridge: Cambridge University Press
- **Schmidt H. A., 2003.** Phylogenetic Trees from Large Datasets Inaugural. Thèse de Doctorat en Mathématique. Dir. Thèse : Von Haeseler A. Univ. Heinrich–Heine–Düsseldorf vorgelegt von Heiko. 123p.
- **Schütze H., 2016.** Les coronavirus dans les organismes aquatiques. *Virologie aquacole*, 327-335.
- **Segondy M., 2020.** Coronavirus humains. *Revue Francophone des Laboratoires*, 526 (1-80) p.
- **Siddell S., Coll M., 1983.** Pathogénese des infections virales. *Clinique vétérinaire de d'Amérique de Nord*, (6) :1049-1079.
- **Stefania L., Camella p., Zecchin B., Zamprin G., Milani A., Tonon F., Georguitti M., Betto M S., Benedictis P., 2020.** Interface entre les chauves- souris et les porcs dans la production porcine lourde. ,13(1) :4.
- **Su S., Gary W., Shi W., Liu.C.J., Lai A.C.K., Zhou J., Liu W., Bi Y., George F., 2016.** Épidémiologie, recombinaison génétique et pathogénèse des coronavirus. *Tendances Microbiol*, 24(6) : 490–502.
- **Takano T., Yamashita S., Michiko OHKUBO M.M., SATOH K., DOKI T., Hohidatsu T., 2016.** Prévalence du coronavirus canin (CCoV) chez le chien au Japon : détection de l'ARN du CCoV et analyse sérologique rétrospective. *The journal of veterinary medical science*, 78(2), 341-345.
- **Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S.,2011.** MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and Parsimony methods [logiciel]. *Molecular Biology and Evolution*.
- **Than V.T., Choe S.E., Vu T.H., 3 Tien D., Nguyen T.L., Bui TH.T., 3 Mai T.N., Cha R.M., Song D., An D.J., 2 et Van P. L., 2020.** Caractérisation génétique du gène

spike des virus de la diarrhée épidémique porcine (PEDV) circulant au Vietnam de 2015 à 2016. *Journal veterinary medicine and science* 6(3) : 535-542.

- **Thompson J. D., Higgins D. G., Gibson T. J., 1994.** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids*
- **Tillier U., Collins RA., 2000.** Les contributions de l'orientation de la réplication, de la direction des gènes et des séquences signal aux asymétries de composition en bases dans les génomes bactériens. *Journal of molecular evolution* 50(3) :249-57. *Research* 22(22): 4673-4680.
- **Tsuda S., Watanabi S., Masangkay J., Alvila P., Ueada N., Iaha K., Taniguchi S., Fuji H., kato K., Horimoto T., Kyowa S., Yuchikawa Y., Akachi H., 2012.** Détection génomique et sérologique du coronavirus de chauve-souris chez les chauves-souris aux Philippines, *Archives de virologie* 157(12) : 2349-55.
- **Vabret A., Dina J., Brison E., Brouard J., Freymutha F.,2009.** Coronavirus humains (HCoV). *Pathol Biol (Paris)*, 57(2): 149–160.
- **Virant M., Černe D., Petrovac M., Paller T., Toplek I., 2021.** Caractérisation génétique et comparaison de trois coronavirus humains (HKU1, OC43, 229E) de patients et de coronavirus bovin (BCoV) de bovins atteints de maladies respiratoires en Slovénie ,13(4) : 676.
- **Vlasova A. N., Saif L.G., 2021.** Le coronavirus bovin et les maladies associées. *Frontiers veterinary science* 8: 643220.
- **Walls A.C., Park Y.J., Tortorici M.A., Wall A., Andrew T. Guire M.C., Velesler D., 2020.** Structure, fonction et antigénicité de la glycoprotéine de pointe SARS-CoV-2. *Revue cell* 181, (281-292) p.
- **Will C., Shinichi N., 2017.** Méthodes comparatives phylogénétiques. *Journal current-biology* 27(9) :333-336.
- **Woo P.C.W., Lau S.K.P., Lam C.C.F., Lau C.C.Y., Tsang A.K.L., Lau J.H., Baï R., Teng J.L.L., Tsang K.C.C., Wang M., Zheng B.J., Chan.K.H., Yuen.K.W., 2012.** Découverte des Sept nouveaux mammifères et aviaires coronavirus dans le genre Deltacoronavirus soutient coronavirus chauve - souris que le gène Source de Alphacoronavirus et Betacoronavirus et coronavirus aviaire comme gène Source de Gammacoronavirus et Deltacoronavirus. *Journal og virology* 86 (7) : 3995-4008.

- **Wrapp D., Wang N., 1, Corbett K.S., Goldsmith J.A., 1Hsieh C.L., Abiona O., Graham S., 2 et McLellan J.S., 2020.** Structure cryo-EM du pic 2019-nCoV dans la conformation de préfusion. 367 (6483).
- **Yin W., Mao C., Luan X., Shan D., Shan Q., Su H., Wong X., Zhou P., Zhao W., Zhang Z., Shen L., Gu X., 2016.** Dynamique évolutive du MERS-CoV : recombinaison potentielle, sélection positive et transmission, 25049.
- **Yuan P., Yang Z., Song H., Wang K., Yang Y., Xie L., Huang S., Liu J., Ran L., Song Z., 2018.** Trois principaux inducteurs de l'infection des entérocytes par l'alphacoronavirus : l'acide sialique, les protéases et un faible pH. Journal de la section de virologie de l'Association internationale des sociétés de microbiologie 61(2) : 53-63.
- **Zhiqiang W., Liang L., Qi J., Jiang D, Li Y., Xianwen R , Bo L., Jinyong J. , Jian Y ., Jie D.,Lilian S., Yafang Z. , Yuhui L., Dandan Z. , Chi Z.,Haoxiang S ., Yuting Z., Hongning Z., Guangjian Z.,Hongying L., Alexei C . Fan Y., Pierre D., Jianwei W., Qiyong L., 2018.** Analyse comparative de viromes de rongeurs et de petits mammifères pour mieux comprendre l'origine faunique des maladies infectieuses émergentes. Journal Microbiome. 6(1) :178.
- **Zhou P., Yang X.L., Wang X.G., Ben H., Zhang L., Zhang W., Si H.R., Zhu Y., Li B., Huang C.L., Chen H.D., Chen J., Luo Y., Guo H., Jiang R.D., Liu M.Q., Chen Y Shen X.R. ,Wang., Zheng X.S., Zhao K., Chen Q.J., Deng F , Liu L ., Yan B., Zhan F.X. , 4, Wang Y.Y., Xiao G.F., ShiUne Z.L., 2020.** Epidémie de pneumonie associée à un nouveau coronavirus d'origine probable de chauve-souris. 579(7798): 270-273. Animale. Dir. Thèse : J-M Elouard. Univ. Lausanne. 111-145. 279p

Biblio-web

Blast: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

FaBox; <https://birc.au.dk/~palle/php/fabox/>

MEGA: <http://www.mybiosoftware.com/mega-5-05-molecular-evolutionary-genetics-analysis.html>

DnaSP: <http://www.ub.edu/dnasp/>

DAMBE : <http://dambe.bio.uottawa.ca/DAMBE/dambe.aspx>

NetWork: https://www.fluxus-engineering.com/sharenet_rn.htm

Fast to rdf: <http://fas2rdf.herokuapp.com/>

Annexe

Animal	<i>Espèce</i>	Région	Gène	Taille en pb	Locus
Chauve-souris	<i>Eidolon helvum</i>	Kenya	RdRp (44)	448	GU065389
	<i>Nycteris thebaica</i>	Mozambique	RdRp (4)	442	MN183196
	<i>Mormopterus jugularis</i>	Madagascar	RdRp (5)	442	MN183187
	<i>Rhinolophus ferrumequinum</i>	Corée du Sud	RdRp (41)	603	MK991935
	<i>Molossus rufus</i>	Brésil	RdRp (9)	426	MH974778
	<i>Molossus molossus</i>		RdRp (2)	426	MH974773
	<i>Eidolon helvum</i>	Arabie Saoudite	RdRp (3)	400	MH396479
	<i>Rhinolophus sp</i>	Inde	Nucléocapside (N) (4)	378	MT656678
	<i>Rhinolophus affinis</i>	Japon	RdRp (10)	440	MK004993
	<i>Hipposideros sp</i>	Zimbabwe	RdRp (7)	415	MG000871
	<i>Rousettus aegyptiacus</i>	Liban	Protéine de nucléocapside N (5)	375	KT346236
	<i>Rousettus aegyptiacus</i>	Égypte	Protéine de nucléocapside N (4)	375	KT346233
	<i>Myotis daubentonii</i>	Danemark	Orf 1a/1b (55)	208	LR025735
	<i>Otomops martiensseni</i>	Kenya	Orf1B (2)	121	GQ920801
	<i>Cardioderma cor</i>	Kenya	Orf1B	121	GQ920802
	<i>Scotoecus sp</i>	Kenya	Orf1B	121	GQ920803
	<i>Tylonycteris pachypus</i>	Chine	S	4059	MW218379
	<i>Pipistrellus pygmaeus</i>	Danemark	Orf1a / 1b (9)	208	LR025739
	<i>Myotis dasycneme</i>	Danemark	Orf 1a/1b (3)	208	LR025731
	<i>Myotis nattereri</i>	Danemark	Orf 1a /1b (2)	208	LR025708
Rat	<i>Eptesicus serotinus</i>	Danemark	Orf 1a/1b	208	LR025705

			(2)		
	<i>Rattus nitidus</i>	Chine	RdRp	433	KY370075
	<i>Lasiopodomys gregalis</i>	Chine	RdRp	440	KY370068
	<i>Apodemus draco</i>	Chine	RdRp	426	KY370072
	<i>Ratus rat</i>	Vietnam	RdRp (11)	407	KX092227
	<i>Hipposideros commersoni</i>	Kenya	Orf1B	121	GQ920806
Canard	<i>Anas sp</i>	Pologne	RdRp (2)	699	KJ690952
	<i>Canard</i>	Chine	Protéine 3a, 3b, E	1171	HM371091
	<i>Anas platyrhynchos</i>	Chine	Replicase 1b (4)	725	KC869937
	<i>Anas platyrhynchos</i>	Chine	Similaire à la réplicase 1b (4)	276	KC869936
	-	Chine	Protéine M	681	JF699749
	<i>Anas platyrhynchos</i>	Chine	Protéine N (5)	549	KC623590
	<i>Anas sp</i>	Pologne	UTR3 (3)	268	KJ690966
	<i>Anas platyrhynchos</i>	Pologne	RdRp (8)	684	KJ690953
Chameau	<i>Camelus dromedarius</i>	Éthiopie	S (2)	4101	MN514969
		Maroc	S (8)	4092	MN514971
		Ethiopie	ORF1ab (4)	2783	MN514979
		Maroc	N protéine	1347	MN514975
Bovin		Chine (10)	S	4092	MN982181
		Corée du Sud	S	4092	HM573326
		Slovénie	RdRp	390	KX059611
Chien	<i>Canis lupus familiaris</i>	Chine	S	4374	LC190907.1
	<i>Canine</i>	Colombie	Membrane protéine (4)	371	MH717724
Civette	<i>Chats civette palmiers</i>	Chine	S (8)	4234	DQ514532

	<i>Civette palmiste</i>	Chine	S (10)	3768	AY627047
Dinde	<i>Meleagris gallopavo</i>	Usa	S	3627	KF652218
Pintade	<i>Pintade</i>	France	S	3951	MG765535
Furet	<i>Chinese ferret</i>	Chine	S	4374	EF192156
	-	Pérou	RdRp (1)	393	KR758768
	<i>Mustela putorius furo</i>	Slovénie	Matrix protéine (4)	321	JX878437
	<i>Mustela putorius furo</i>	Slovénie	Nucleo-capsid protéine (4)	321	JX878441
Humain	<i>Homo sapiens</i>	France	S	4077	KF963240
	<i>Homo sapiens</i>	Pays-Bas	S	4071	MN488637
	<i>Homo sapiens</i>	Chine	S	4071	DQ437613
	<i>Homo sapiens</i>	Arabie Saoudite	S	4062	KT806016
	<i>Homo sapiens</i>	USA	S	3822	MZ027646
	<i>Homo sapiens</i>	Iran	S	3822	MW136267
	<i>Homo sapiens</i>	France	Nucléoprotéine (20)	1134	KY862074
	<i>Homo sapiens</i>	Pays Bas	nucleocapsid protéine (4)	1134	MN488631
	<i>Homo sapiens</i>	Italie	ORF1ab (2)	423	EU030685
	<i>Homo sapiens</i>	Slovénie	RdRp (66)	390	KX059697
	<i>Homo sapiens</i>	Norvège	ORF1b (15)	208	AM261824
	<i>Homo sapiens</i>	Brésil	RdRp (20)	404	JN251802
	<i>Homo sapiens</i>	Chine	ORF1a / b (20)	395	KP347657
Souris	<i>Myodes rufocanus</i>	Chine	RdRp	439	KY370066
	<i>Mus caroli</i>	Chine	RdRp (2)	440	KY370058
	<i>Apodemus sylvaticus</i>	Chine	RdRp	428	KY370067
Porc	Cochon	Chine	S	4344	MH167923
	Porcine	Canada	S	4271	KM189366
	Cochon	Mexique	S	4200	KY828999

	Cochon	Chine	S	4161	MN725088
	Porcine	Chine	S	4161	MT090140
	Porcine	Japon	S	4161	KY619734
	<i>Sus scrofa</i>	Chine	S	4161	MN508818
	Porcine	Thaïlande	S	4161	KX981900
	Porcine	Vietnam	S	4158	KX708901
	Porcine	Espagne	S	3678	M94097.1
	Cochon de nourrice	Chine	S	3480	KX534090
	Porcine	USA	S	3483	KP995356
	<i>Sus scrofa</i>	Mexique	S	3483	MK478381
	Swine	Italie	RdRp	396	MW089343
	Cochon	Chine	Orf1a (10)	1249	MN173801
	Cochon	Chine	N protéine (10)	1029	MN173792
Yak	Yak	China	S protéine (20)	4092	MH810162
Poulet	Poulet	Chine	S	3480	MK248485
	Poulet	Chine	1ab (2)	426	KP033136
Chat	<i>Felis catus</i>	Chine	S	4359	MK987175
Chat	<i>Chat domestique</i>	Chine	S	1992	MT181985
	<i>leopard cat</i>	China	protéine 7b (5)	931	MT112942
	<i>Feline</i>	Colombie	Protéine membrane	371	MH717723
	<i>Felis catus</i>	Liban	ORF1ab poly-protéine (29)	369	MW899028
	<i>Spotted hyena</i>	Tanzanie	Nucleocapsid (9)	1149	MW505904