Original Research Paper

A study of genetic variants of SARS-CoV-2 using bioinformatics tools.

Derouiche L.[1]*, Benzayed Y.[2], Belmihoub M.[2], Derouiche F.[3]

[1]*Higher School of Food Sciences & Agrifood Industries, Algiers, Algeria.*
[2]*Saad Dahlab University Blida, Blida, Algeria.*
[3]*Faculty of Medicine, University of Algiers - Benyoucef Benkhedda. Algeria.*

Corresponding Author: Derouiche Louiza, Higher School of Food Sciences & Agrifood Industries, Algiers, Algeria. **E-mail**: derouiche_fatma@yahoo.fr

Abstract

The severe acute respiratory syndrome coronavirus 2, that is commonly known as SARS-CoV-2, appeared for the first time in December 2019 in the City of Wuhan (China) and has since affected most countries around the world, hence becoming a major global threat to all humans. The present study was carried out for the purpose of better understanding the molecular structure of this virus. It is based on the inventory and processing of all DNA sequences that have been published to date on the general biological data storage bank GenBank. In order to carry out this study, it was deemed necessary to use various bioinformatics tools such as MEGA 11 software for building the phylogenetic trees, DnaSP 6 for identifying haplotypes, NetWork 10 for determining phylogenetic networks, and DAMBE 7 to perform statistical analyses which were then supplemented by the new DnBA program that was developed in the present work. These analyses allowed us to classify the 11 SARS-CoV-2 genes under study into 3three categories; first, the most variable genes, such as ORF1ab, ORF3a, N, and S, next, the less variable genes, such as ORF8, ORF10, ORF7a, and ORF7b, and then the unvaried genes, such as ORF6, E, and M.

Keywords: SARS-CoV-2; DNA; GenBank; Bioinformatics; Genes.

Introduction

SARS-CoV-2 belongs to the Coronaviridae family. These are unsegmented single-stranded RNA viruses, 26 to 32 kilobases in size (Zi-Wei and Shuofeng, 2020; Patrick et *al.,* 2020).

Viruses of the Coronaviridae family are named after their appearance. Indeed, when examined under an electron microscope, they exhibit a crownlike, or coronal, appearance. Based on the above findings, on February 11, 2020, the International Committee for the Classification of Viruses gave the name Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) to this new coronavirus which was first identified in December 2019 in Wuhan in the Chinese province of Hubei (Sukhadeo et *al.,* 2020; Wang et *al.,* 2021). It is important to know that SARS-CoV-2 is classified as the third Coronavirus, after MERS Covid and SARS Covid. Note that this virus is highly pathogenic and can easily infect humans (Huihui et *al.,* 2020).

The definitive diagnosis of SARS-CoV-2 is based on the molecular identification of the virus by the RT-PCR test that is performed on respiratory samples. However, considering the difficult and insufficient accessibility to this technique, and the high false-negative rate

given by this technique (around 30%), the diagnosis of SARS-CoV-2 can be made using the combination of suggestive clinical signs in addition to a compatible scenographic image, such as the thoracic tomodensitometric examination (Hoppenot, 2020; Plaçais and Richier, 2020; Jawerth, 2020). This diagnosis can also be performed by means of a number of immunological tests (Gala et *al.,* 2020; Mahalaxmi et *al.,* 2020). The results of the RT-PCR technique, obtained in the various molecular biology research laboratories around the world, are stored in databases in order to allow all researchers in the field to use them in order to achieve research that may help in humanity's battle against the pandemic caused by this virus. For this reason, it was decided to conduct this study which is based on the inventory of all DNA versions of SARS-CoV-2 sequences published to date. Then, we proceeded to the processing and classification of these sequences using bioinformatics tools.

Material and methods

The present work aims at discussing the sequences obtained by the Reverse Transcriptase - Polymerase Chain Reaction (RT-PCR) test which is the reference technique for the diagnosis of SARS-CoV-2. In this regard, Plaçais and Richier (2020) as well as Jawerth (2020) reported that the diagnosis of SARS-CoV-2 using the RT-PCR technique can be carried out in several steps. The first step is to take and treat samples in order to extract only the RNA; then, the RNA is converted into DNA; afterwards, certain well-targeted fragments of the DNA are then amplified and sequenced. It is worth indicating that the sequences obtained by different research laboratories in different countries around the world are in most cases submitted to various databases, such as GenBank which is a storage bank for generalist sequences and biological data. It should be mentioned that the GenBank sequence database is an open access, recorded collection of freely available nucleotide sequences. All researchers around the world can voluntarily consult the data stored in GenBank (Sayers et *al.,* 2020).

First, we extracted all SARS-CoV-2 DNA sequences published on GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Subsequently, we used several DNA sequence processing computer programs to analyze and classify the different sequences obtained. It should be noted that we used the latest versions of a series of specialized DNA molecular data processing programs that were either downloaded or used online (FaBox, MEGA, DnaSP, DAMBE, NetWork). In addition, a new computer program was developed to perform more advanced data analysis and processing (DnBA).

-FaBox

FaBox (Toolbox for Fasta sequences) is a collection of simple and intuitive web services (https://users-birc.au.dk/palle/php/fabox/) that biologists and medical researchers can use to quickly carry out a typical task using sequence data. These web services help us immensely in extracting, modifying and replacing sequence headers; they also allow us to join or split datasets based on header information. There are also other services involving the fusion of a set of sequences into haplotypes and the automated formatting of input files for a number of population genetics programs (Villesen, 2007). It is worth noting that the use of FaBox (version 1.5) is a preliminary and mandatory step for the preparation of DNA sequences in order to process and analyze them based on the different programs used in the rest of our work

-MEGA

Molecular Evolutionary Genetics Analysis (MEGA) is a computer program for performing statistical analysis of molecular evolution and building phylogenetic trees. It was developed to provide a series of integrated tools for biologists to conduct statistical DNA analyses and to process protein sequence data from an evolutionary perspective. This software allows the construction of sequence alignments and phylogenetic trees. In addition, evolutionary bioinformatics methods in biology, biomedicine and evolution were also used. It should be mentioned that the data analysis methods are generally selected by the user (Tamura, 2011). Moreover, the MEGA 11 version was used to carry out the phylogenetic analysis, and to obtain the phylogenetic trees after several processing and analysis steps which consist in introducing the DNA sequences in FASTA form, next aligning these sequences using Clustal W,

the, choosing the method for building the phylogenetic tree and finally adjusting the parameters of the selected method which, in the present case, is the unweighted pair group method with arithmetic mean (UPGMA) method. This simple method allows the transformation of a matrix of distances (between different organisms, populations, or nucleotide sequences) into a rooted tree based on similarities between pairs of sequences (Yoann, 2012) with a Bootstrap of 1000.

-DnaSP

DnaSP (DNA Sequence Polymorphism Analysis of Large Data Sets) is a bioinformatics tool designed for comprehensive variation analysis of DNA sequence data. This program is widely used in comprehensive population genetic analyses on multiple sequence alignments (Rozas et *al.,* 2017). DnaSP 6.12.03 is a version which incorporates new features that are particularly suitable. This version makes it possible to determine and verify the nucleotide composition using haplotypes.

-DAMBE

DAMBE (Data Analysis in Molecular Biology and Evolution) is a computer program that is generally used for descriptive and comparative analysis of molecular data. This integrated software is particularly used in the conversion, manipulation, statistical and graphical description, as well as the analysis of nucleotide sequence data (Xia, 2018). Note that the use of DAMBE version 7.3.2 makes it possible to conduct statistical studies on nucleotide bases such as adenine (A), cytosine (C), guanine (G) and thymine (T).

-DnBA

We designed and developed the DnBA (DNA Basis Analysis) program in order to integrate more information in order to support the results obtained by the DAMBE software by calculating the percentages of nitrogen bases (A, T, C, G) for each DNA sequence studied, as well as the ranges of percentages of the nitrogenous bases (A, T, C, G) for all DNA sequences of the same gene.

-NetWork

NetWork software is used to reconstruct phylogenetic networks, and to deduce ancestral and potential types, ramifications and evolutionary variants, and also to assess the dating. It is worth indicating that the network generates evolutionary trees and networks from genetic, linguistic and other data (Huson and Bryant, 2012). A very recent version of NetWork software, i.e. NetWork 10.2.0.0, was used in SARS-CoV-2 research as it allows users to easily create rdf SARS-CoV-2 files (Forster and Forster, 2020) because this file form can be easily and directly recognized by NetWork when building phylogenetic networks.

Results

GenBank

The processing of 498 626 DNA sequences of SARS-CoV-2, deposited in the GenBank database, allowed us to classify 11 genes which represent the most studied genes by different research laboratories in 45 countries. These are either structural genes M (Membrane protein), N (Nucleocapsid protein), S (Spike glycoprotein), E (Envelope protein)), or functional genes (Open Reading Frames (ORF): ORF1ab (ORF1a and ORF1b), ORF3a, ORF7a, ORF7b, ORF6, ORF8, ORF10) (Rozhgar et *al.,* 2020; Sekulic et *al.,* 2020; Juckel et *al.,* 2020). The results obtained are reported in Table I.

Table I. SARS-CoV-2 genes with their nitrogen base pair (bp) size intervals, and the different countries of origin of the DNA samples analyzed.

Genes	Sizes in bp	Country
ORF1ab	[21100-21400]	China, USA, Spain, Israel, Nepal, Korea, Brazil, Iran, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Germany, Turkey, Serbia, Greece, Italy, Egypt, Tunisia, India, Netherlands , Sweden, Chile, Sri Lanka,

		Kazakhstan.
S	[3800-3900]	China, USA, Kenya, Spain, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Italy, Iran, Sweden, Sri Lanka, Kazakhstan , Netherlands, Chile, Egypt, Tunisia, India.
ORF3a	[800-900]	China, USA, Kenya, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Peru, France, Bangladesh, Czech, Germany, Serbia, Greece, Italy, Egypt, Tunisia, India, Sweden, Iran, Sri Lanka , Kazakhstan, Netherlands, Chile.
E	[200-300]	China, USA, Kenya, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Sweden Italy, Egypt, Tunisia, India.
M	[600-700]	China, USA, Kenya, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Iran, Sri Lanka, Kazakhstan, Country bottom, Chile, Sweden, Italy, Egypt, Tunisia, India.
ORF6	[150-200]	China, USA, Kenya, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Italy, Egypt, Tunisia, India.
ORF7a	[300-400]	China, USA,,, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Italy, Iran, Sri Lanka, Kazakhstan, Netherlands, Chile, Sweden, Egypt, Tunisia, India.
ORF7b	[100-150]	China, USA, Spain, Israel, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Iran, Sri Lanka, Kazakhstan, Netherlands, Chile, Egypt, Tunisia.
ORF8	[350-400]	China, USA, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Iran, Sri Lanka, Kazakhstan, Netherlands, Chile, Italy, Egypt, Tunisia, India, Sweden.
N	[1000-1500]	China, USA, Kenya, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, India, Italy, Egypt, Tunisia. Iran, Sri Lanka, Kazakhstan, Netherlands, Chile, Sweden.
ORF10	[90-140]	China, USA, Kenya, Spain, Israel, Colombia, Korea, Brazil, Taiwan, Vietnam, Finland, Pakistan, Peru, France, Japan, Bangladesh, Czech, Germany, Turkey, Serbia, Greece, Italy, Egypt, Tunisia, India, Iran, Sri Lanka, Kazakhstan, Netherlands, Chile, Sweden.

Phylogenetic analysis

MEGA software allowed us to build 11 phylogenetic trees for 11 different genes (M, N, S, E, ORF1ab, ORF3a, ORF7a, ORF7b, ORF6, ORF8, and ORF10). Each gene is represented by several DNA sequences, depending on the country of origin of the samples. Table II gives an overview of the structure of the different trees, with the exact number of branches, nodes and leaves for each phylogenetic tree obtained.

Table II. Summary of the results obtained for the phylogenetic trees of the SARS-CoV-2 genes obtained by MEGA software.

Root	Branches	Nodes	Leaves
ORF1ab	30	3	10
S	15	3	25
ORF3a	8	3	36
E	43	0	0
M	43	0	0

ORF6	41	0	0
ORF7a	39	0	0
ORF7b	27	2	26
ORF8	3	2	25
N	16	3	30
ORF10	3	1	37

Close examination of all phylogenetic trees revealed that there are two types of trees. The first type is very rich in ramifications and includes the genes S, N, ORF1ab (Figure 2), ORF3a; while, the second type is poor in ramifications and comprising the genes M, E, ORF7a, ORF7b, ORF6 (Figure 2), ORF8, and ORF10.

Note: It is useful to bear in mind that for some countries, like China, there are several research laboratories, in different regions, which are currently working on SARS-CoV-2. Therefore, in order to study the variations of genes within the same country, it was deemed interesting to use all the DNA sequences by adding numbers to the name of the country (Example: China 1, China 2).

Haplotypes and haplotype diversity

The results obtained for the SARS-CoV-2 genes under study by the DnaSP software are organized into groups, in the form of haplotypes, with significant positive haplotype diversity. These haplotypes include the genes: ORF1ab, ORF3a, ORF10, N, ORF8, S, ORF7a, and ORF7b, which are presented in Table III. These groups can also be in the form of a single haplotype (H = 1) and haplotype diversity equal to zero (Hd = 0.000), as is the case for genes E, M and ORF6.

Statistical analysis

The use of DAMBE and DnBA allowed us to perform additional molecular statistical calculations and studies. The results obtained are divided into two groups:

- The first group comprises the sequences of E, M and ORF6 genes with identical percentages of A, T, C and G, for all the studied sequences of the same gene obtained from different countries.

- The second group includes the sequences of S, N, ORF1ab, ORF3a, ORF7a, ORF7b, ORF10 and ORF8 genes with different percentages of A, T, C and G, for the studied sequences of the same gene obtained from different countries.

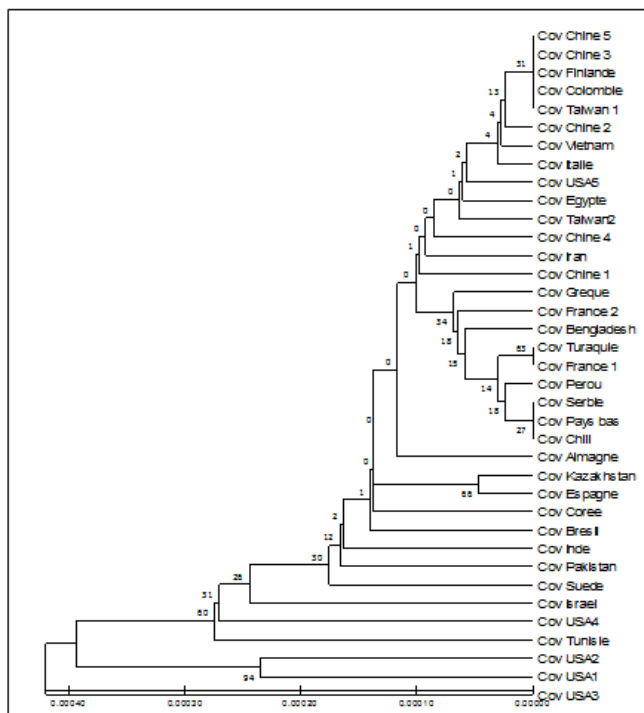Figures 3, 4, 5 and 6 present some results obtained by the two programs DAMBE and DnBA

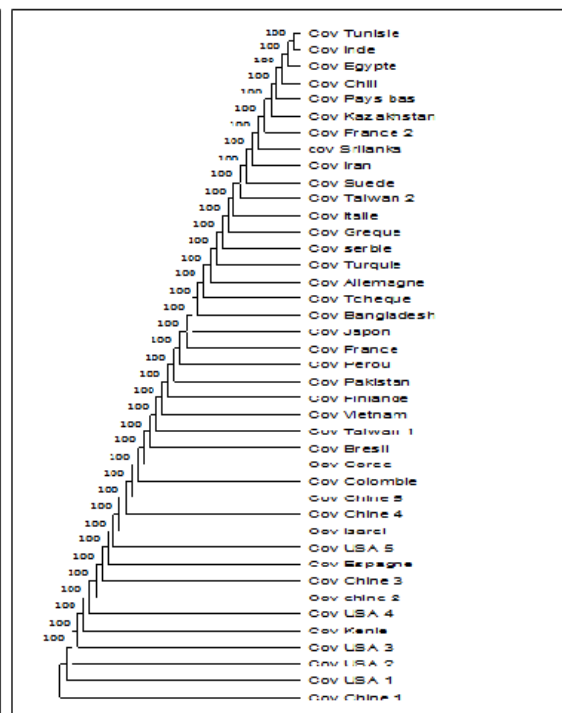**Figure 1.** Phylogenetic tree of the ORF1ab gene constructed from 37 DNA sequences.

Figure 2. Phylogenetic tree of the ORF6 gene constructed from 41 DNA sequences.

Table III. The results of analysis of SARS-CoV-2 genes obtained by the DnaSP software.

Genes	Number of sequences used	Total number of sites	Number of variable sites	Number of haplotypes (H)	Haplotypic diversity (HD)
ORF1ab	37	21281	146	30	0.9790
ORF3a	41	21255	98	29	0.9778
ORF10	39	117	2	3	0.1012
N	43	1260	23	16	0.7431
ORF8	26	120	1	2	0,0769
S	37	3748	24	15	0.7867
ORF7b	26	120	1	2	0,0769
ORF7a	39	366	1	2	0,0513

```
Chine1:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
USA1:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
USA2:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
USA3:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
USA4:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Chine2:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
chine3:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Espagne:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
USA5:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Israel:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Chine4:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Chine5:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Colombie:    A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Coree:       A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Brezil:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Taiwan 1:    A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Vietnam:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Finlande:    A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Colombie2:   A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Perou:       A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
France1:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Japon:       A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Bangladch:   A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Tcheque:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Almagne:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Serbie:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Greque:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Italie:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Taiwan2:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Suede:       A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Iran:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Srilanka:    A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
France2:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Kazakhstan:  A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Paysbas:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Chili:       A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Egypte:      A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Tunisie:     A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb
Inde:        A: 25.561%  C: 21.824%  G: 20.777%  T: 31.839%  (A-T/A+T): -0.025  (G-C/C+G): -0.109  Total: 669pb

        A: 25.561%-25.561%   C: 21.824%-21.824%   G: 20.777%-20.777%   T: 31.839%-31.839%

min (G-C/C+G)= -0.025 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie,Coree,Brezil,Taiwan,Vietnam,Finlande,Colombie2,
max (G-C/C+G)= -0.025 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie,Coree,Brezil,Taiwan,Vietnam,Finlande,Colombie2,
min (A-T/A+T)= -0.109 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie,Coree,Brezil,Taiwan,Vietnam,Finlande,Colombie2,
max (A-T/A+T)= -0.109 avec 39 pays: Chine1,USA1,USA2,USA3,USA4,Chine2,Chine3,Espagne,USA5,Israel,Chine4,Chine5,Colombie,Coree,Brezil,Taiwan,Vietnam,Finlande,Colombie2,
```

Figure 3. Results of nitrogen base analysis of M gene sequences obtained by DnBA.

```
Chine1:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
USA1:        A: 27.174%  C: 20.894%  G: 18.478%  T: 33.454%  (A-T/A+T): -0.061  (G-C/C+G): -0.104  Total: 828pb
USA2:        A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
USA3:        A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Kenie:       A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Chine2:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Chine3:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
USA5:        A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Israel:      A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Chine4:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Chine5:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Colombie:    A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Coree:       A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Brazil:      A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Taiwan:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Vietnam:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Finland:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Colombie2:   A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Perou:       A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
France:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
France2:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Bengladech:  A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Tcheque:     A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Almagne:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Serbie:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Greque:      A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Italie2:     A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Taiwan2:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Suede:       A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Iran:        A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Srilanka:    A: 27.174%  C: 21.014%  G: 18.599%  T: 33.213%  (A-T/A+T): -0.061  (G-C/C+G): -0.100  Total: 828pb
France3:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Kazakhstan:  A: 27.174%  C: 21.014%  G: 18.357%  T: 33.454%  (A-T/A+T): -0.067  (G-C/C+G): -0.104  Total: 828pb
Paysbas:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Chili:       A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Egypte:      A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Tunisie:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Inde:        A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb
Italie3:     A: 27.174%  C: 21.014%  G: 18.478%  T: 33.333%  (A-T/A+T): -0.064  (G-C/C+G): -0.102  Total: 828pb

        A: 27.174%-27.174%   C: 20.894%-21.014%   G: 18.357%-18.599%   T: 33.213%-33.454%

min (G-C/C+G)= -0.067 avec 7 pays: Israel,Coree,Brazil,Tcheque,Italie2,Suede,Kazakhstan
max (G-C/C+G)= -0.061 avec 1 pays: Srilanka
min (A-T/A+T)= -0.104 avec 8 pays: USA1,Israel,Coree,Brazil,Tcheque,Italie2,Suede,Kazakhstan
max (A-T/A+T)= -0.100 avec 1 pays: Srilanka
```

Figure 4. Results of nitrogen base analysis of ORF3a gene sequences obtained by DnBA.

Figure 5. Identical sequences of the ORF1ab gene obtained by DAMBE.

Figure 6. Results of the statistical analysis of an ORF1ab gene sequence obtained by DAMBE.

Construction of phylogenetic networks

The networks generated by the NetWork program represent the evolutionary and phylogenetic relationships that exist between the different DNA sequences of the SARS-CoV-2 genes analyzed. The results obtained are distributed among three groups as expressed below:
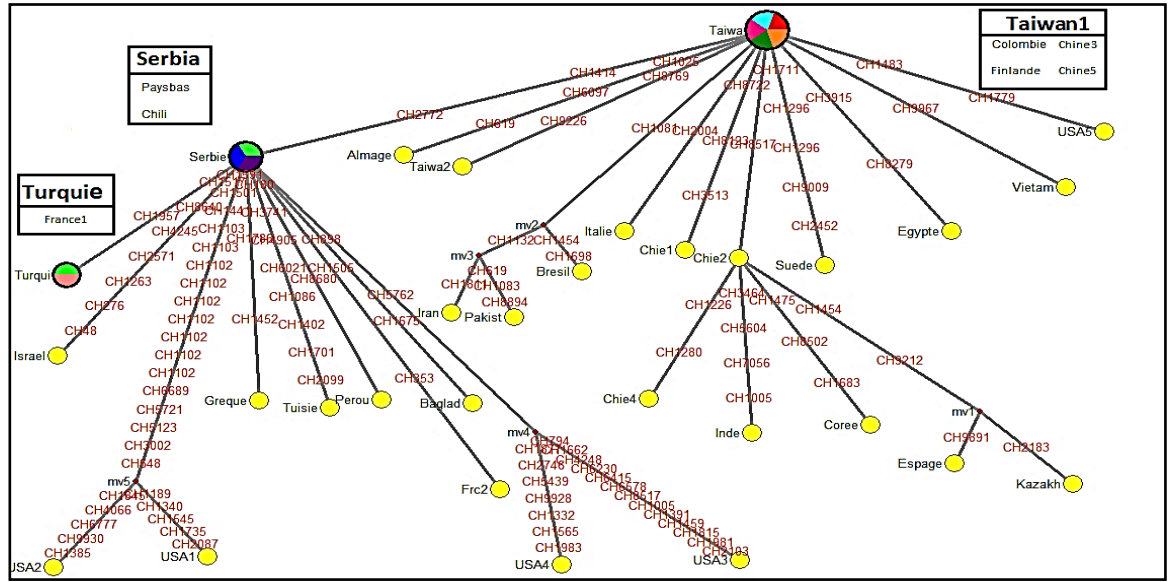
Figure 7. Phylogenetic network of ORF1ab genes, constructed from 37 DNA sequences of different origins, using NetWork.

- The first group is the one for which we could build a specific phylogenetic network for each gene. This is the case for N, ORF1ab (Figure 7), ORF3a, ORF10 and S.

- The second group is the one for which we were unable to build the phylogenetic networks, despite the large number of DNA sequences studied for each gene, though all the sequences were practically

identical. It should be noted that each gene in this group contains less than 3 different sequences; this is the case for genes: ORF7a, ORF7b and ORF8.

- The third group is the one for which we were unable to build the phylogenetic networks because all the DNA sequences studied are identical for each gene; this is the case for the genes E, M and ORF6.

Discussion

Computer programs, like MEGA 11, DnaSP 6, DnBA, DAMBE 7 and NetWork 10, were used for the analysis of SARS-CoV-2 genes. It turned out that these genes can be divided into 3 groups:

The first group concerns the most variable genes, namely S, N, ORF1ab and ORF3a, which are represented by phylogenetic trees that are rich in ramifications; they were obtained by the UPGM method. According to Mercatelli and Giorgi (2020), these genes are among the most susceptible to mutations. At the same time, their phylogenetic networks, which were generated by the software NetWork, suggest that these genes are round in shape (circles), either with a single color (yellow) signifying a single sequence, or with several colors implying the presence of several sequences of different origins. It is worth indicating that these circles represent nodes in phylogenetic trees that bind together with branches which represent the genetic distances that separate them. These results are similar to those reported in the studies conducted by Ridley and Rasmont (1997), Gattolliat (2002), and Schmidt (2003) who all defined the distance between two nodes as a branch. It is interesting to mention that the branches can be evaluated, which means that they can be measured (i.e. a distance, an evolution rate, a number of mutations). This distance, which suggests the presence of mutations, is reflected in the detection of several haplotypes of the same gene; this is defined as the haplotype diversity (Hd) that can be detected using a computer program like for example DnaSp (Anderson and Karel, 2011).

The second group includes the least variable genes, namely ORF8, ORF10, ORF7a and ORF7b, which exhibit haplotype diversity (Hd), calculated by DnaSp 6, with a value between 0.0513 and 0.1012; whereas the most variable genes possess a haplotype diversity value between 0.7431 and 0.9790. Xia (2000) suggested that this high haplotype diversity is due to the high rate of transition or transversion mutations. On the other hand, Roy et *al.* (2020) revealed that SARS-CoV-2 genes may undergo transitions and transversions, which justifies their haplotype diversity.

The third group includes unvaried genes, such as M, E and ORF6, for which all the DNA sequences representing one gene are identical, which means that they do not exhibit any haplotype diversity (Hd = 0) because they possess a unique haplotype that is expressed by simple phylogenetic trees, with identical genetic distances. Just like the least variable genes, unvaried genes do not exhibit haplotype networks because they contain less than 3 different sequences, except for the ORF10 gene which has a phylogenetic network. This is represented as a phylogenetic tree that contains 3 branches and consequently 3 different groups of sequences (39 representative sequences).

Furthermore, the DnBA and DAMBE programs allowed us to perform several statistical treatments for the purpose of comparing the percentages of bases A, T, C and G, for the different sequences for each gene obtained from different countries. The results obtained indicated that the genes S, N, ORF3a and ORF1ab were among the most variable ones. According to Mercatelli and Giorgi (2020), these genes can undergo mutations. These same researchers indicated that there are differences in the percentages of the bases A, T, C and G between the DNA sequences of each gene. With regard to Xia (2013), and Xia and Xie (2001), they found out that the variations in nitrogenous bases mean that mutations can occur. As for the least variable genes, such as ORF8, ORF10, ORF7a and ORF7b, it was observed that the percentages of the bases A, T, C and G between the DNA sequences of each gene were close to each other but not identical. However, these percentages were similar for the genes E, M and ORF6 which are unvaried genes. Likewise, Koyama et al. (2020) proved that variations or mutations of C, G, A and T mean that the sequence has undergone transitions and transversions. The above findings

allowed us to determine the haplotypes as well as the haplotype diversity for genes that exhibited variations.

It is worth indicating that gene S encoding the Spike glycoprotein was found among the genes that were classified in the group of most variable genes. In this context, Philipe (2000) indicated that the Spike glycoprotein gives crown-like shapes on the surface of the virus that binds to the receptor in the lung tissue. On account of its structure and functional characteristics, this protein can be used as an antigen in vaccination strategies (Sternberg, 2020). It contains antigenic sites that are exposed on the virus surface and are therefore accessible to the immune system, thus constituting antigens that can be recognized by the antibodies produced by infected hosts (Ni et *al.,* 2020). However, it should be noted that the genomic sequence encoding this protein, which is potentially antigenic, exhibits great variability between various viral species. This variability results from the selection of genetic mutations, which allows the viruses to evade the immune response of the host (Sallard et *al.,* 2020). These findings justify the high rate of variability that was detected for this gene.

Furthermore, the results of the present work showed that gene N, just like gene S, can be classified in the group of the most variable genes. It is worth bearing in mind that several other works have previously suggested that non-synonymous mutations developed in protein S as the pandemic SARS-CoV-2 progressed. Conversely, Dutta et *al.* (2020) found out that the gene N is more conserved and more stable, with 90% amino acid homology and fewer mutations over time.

In addition, the SARS-CoV-2 genome also contains a variable number of open reading frames (ORFs). According to the work of Jean Michel et *al.* (2020), the different ORFs encoding accessory proteins, which are not essential for virus replication, appear to play a role in pathogenesis. These findings are in good agreement with those reported by Finkel et *al.* (2020) who suggested that ORFs are not conserved in most Sarbecovirus 19 (a subgenus of SARS-CoV-2). With regard to Hassan et *al.* (2021), they reported rare mutations in the accessory proteins ORF6, ORF7b and ORF10 of the SARS-CoV-2 genomes. These results confirm the structure and variability of the ORF sequences previously investigated and discussed.

Moreover, it was revealed that the M and E genes are non-variable genes. In this context, Thomas (2020) and Bianchi et *al.* (2020) indicated that the conservation of the DNA of M and E genes is due to the cooperation between M and E proteins and S protein, which means that any mutation in DNA can contribute to attachment to the host cell, and facilitate the entry of viruses.

Conclusion

No one denies the fact that the coronavirus pandemic is an event that has marked, and even upset, the entire world during this century. It has indeed affected more than two hundred million people to date, and has killed at least four million people around the world, not to mention its immense and catastrophic impact on the global economy. To end this pandemic, huge amount of information is continually being provided about this virus by scientific research laboratories around the world, and is then submitted to databases like GenBank. This study was conducted using the data from GenBank which allowed us to inventory SARS-CoV-2 sequenced genes around the world. In addition, these genes culd be analyzed using different bioinformatics programs, like MEGA 11, DnaSP 6, NetWork 10, DAMBE 7 and DnBA, which helped us in identifying three categories of genes. The first category includes the most variable genes (ORF1ab, ORF3a, N, S), the second category comprises the least variable genes (ORF8, ORF10, ORF7a, ORF7b) and the third category contains the unvaried genes (ORF6, E, M). It should also be noted that the most variable and least variable genes are less conserved due to transitions and transversions, which is reflected in the number of haplotypes.

References

Anderson J. D. Karel WJ. 2011. A Genetic Assessment of Current Management Strategies for Spotted Seatrout in Texas, Marine and Coastal Fisheries, 1:1, p: 121-132.

Bianchi M. Benvenuto D. Giovanetti M. Angeletti S. Ciccozzi M. Pascarella S. 2020. Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics BioMed Research International, ID 4389089, 6 p.

Dutta NK. Mazumdar K. Gordyc JT. 2020. The Nucleocapsid Protein of SARS–CoV-2: a Target for Vaccine Development. J. Virol. 94: e00647-20.

Finkel Y. Mizrahi O. Nachshon A. Weingarten-Gabbay S. Morgenstern D. Yahalom-Ronen Y. Tamir H. Achdout H. Stein D. Israeli O. Beth-Din A. S Melamed S. Weiss S. Israely T. Paran N. Schwartz M. Stern-Ginossar N. 2020. The coding capacity of SARS-CoV-2. Nature, Vol. 589, 125 p.

Forster P. Forster M. 2020. Network Release notes. (https://www.fluxus-engineering.com/sharenet_rn.htm)

Jawerth N. 2020. Comment la RT-PCR en temps réel permet-elle de détecter le virus de la COVID-19 ?. IAEA: L'Agence internationale de l'énergie atomique. Maladies infectieuses. Vol. 61-2.

Juckel D. Dubuisson J. Belouzard S 2020. Les coronavirus, ennemis incertains, Coronavirus, emerging viruses. Med Sci (Paris), Vol. 36, Numb. 6-7, p : 633-641.

Hassan SkS. Choudhury PP. Roy B. 2021. Rare mutations in the accessory proteins ORF6, ORF7b, and ORF10 of the SARS-CoV-2 genomes. Meta Gene. 28: 100873.

Hoppenot I. 2020. Imagerie pulmonaire de la Covid-19 : Le scanner, un examen clé à prescrire à bon escient. VIDAL. L'intelligence médicale au service du soin.

Huihui W. Xuemei L. Tao L. Shubing Z. Lianzi W. Xian W. Jiaqing L. 2020. The genetic sequence, origin, and diagnosis of SARS-CoV-2. Eur. J. Clin. Microbiol. Infect. Dis. p: 1-7.

Huson DH. Bryant D. 2012. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol, 23(2) p: 254-67.

Gala JL. Nyabi O. François Durant J. Chibani N. Bentahir M 2020. Méthodes diagnostiques du COVID-19. Biologie Et Technologies Moléculaires Appliquées. p: 228-235.

Gattolliat JL 2002. Etude systématique, cladistique et biogéographique des Baetidae (Ephemeroptera) de Madagascar. Thèse de doctorat en Zoologie et Ecologie Animale. Dir. Thèse : J-M Elouard. Univ. Lausanne. 111-145. 279p

Koyama T. Platt D. Parida L. 2020. Variant analysis of SARS-CoV-2 genomes. Bull World Health Organ 1, 98(7) p: 495-504.

Mahalaxmi L. Kaavya J. Mohana DS. Soo B. L. Ahmed AD. Ssang-goo C. Balachandar V 2020. COVID-19: an update on diagnostic and therapeutic approaches. BMB, 53(4) p:191-206.

Michel CJ. Mayer C. Poch O. Thompson JD. 2020. Caractérisation des gènes accessoires dans les génomes du coronavirus. Revue de virologie, Vol. 17, Numb. 131.

Mercatelli D. Giorgi FM 2020. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front. Microbiol. 11:1800.

Ni L. Ye F. Cheng ML. Yu Feng Y. Deng Y. Q. Zhao H. Wei P. Ge J. Gou M. Li X. Sun L. Cao T. Wang P. Zhou C. Zhang R. Liang P. Guo H. Wang X. Qin CF. Chen F. Dong C. 2020. Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. Immunity 16, 52(6) p: 971-977.

Patrick CY. Woo Susanna KP. Lau LSFC. Lau CY. Tsang AKLC. Lau JHN. Bai R. Teng LLJ. Tsang CCC. Wang M. Zheng B. Chan K. Yuen K 2020. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. Journal of Virology, Vol. 86, Numb. 7 p: 3995-4008.

Philippe S 2021. 11 avril 2020-Infection à SARS-Cov 2 : ce que doit savoir l'hépato-gastroentérologue. Hôpital Saint-Antoine, Service de Gastroentérologie et Nutrition, 184 rue du Faubourg Saint-Antoine, 75571 Paris cedex 12.

Plaçais L. Richier Q 2020. COVID-19: Clinical biological and radiological characteristics in adults, infants and pregnant women. An up-to-date review at the heart of the pandemic. Rev Med Interne, 41(5) p: 308-318.

Ridley M. Rasmont R 1997. Evolution biologique. Paris, Bruxelles, De Boeck Université 719 p.

Roy C. Mandal SM. Mondal SK. Mukherjee S. Mapder T. Ghosh W. Chakraborty R 2020. Trends of mutation accumulation across global SARS-CoV-2 genomes: Implications for the evolution of the novel Coronavirus. Genomics, 112(6) p: 5331-5342.

Rozas J. Ferrer-Mata A. Sanchez-DelBarrio J. Guirao-Rico S. Librado P. Ramos-Onsins S. Sanchez-Gracia 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol. Biol. Evol, 34(12) p: 3299-3302.

Rozhgar AK. Safdarb M. Ozaslan M 2020. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 19:100682.

Sallard E. Halloy J. Casane D. Helden J-V. Decroly E 2020. Retrouver les origines du SARS-CoV-2 dans les phylogénies du coronavirus. Med Sci (Paris), Vol. 36, Numb. 8-9, p: 783-796.

Sayers E. Cavanaugh M Clark K. Pruitt K. Schoch C. Sherry S. Karsch-Mizrachi I 2020. GenBank. Nucleic Acids Research, 49(10). 1093p.

Schmidt HA 2003. Phylogenetic Trees from Large Datasets Inaugural. Thèse de Doctorat en Mathématique. Dir. Université de Heinrich–Heine–Düsseldorf vorgelegt von Heiko, 123p.

Sekulic M. Harper H. Nezami BG. Shen L. Sekulic SP. Koeth AT. Harding VC. Gilmore H. Sadri N 2020. Molecular Detection of SARS-CoV-2 Infection in FFPE Samples and Histopathologic Findings in Fatal SARS-CoV-2 Cases. Am. J. Clin. Pathol.: aqaa091.

Sternberg A. Naujokat C 2020. Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination. Life Sci. 15. 257p.

Sukhadeo BB Deepak BR. Pankaj PG. Jess V 2020. Global scenario, public health concerns and mitigation strategies to counter current ongoing SARS-CoV-2 / COVID-19 pandemic. Hum Vaccin Immunother. p: 1-11.

Tamura K. Peterson D. Peterson N. Stecher G. Nei M. Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28(10) p: 2731-2739.

Thomas S 2020. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. Pathogens and Immunity. 5(1) p: 342-363.

Villesen 2007. FaBox: an online toolbox for FASTA sequences. Molecular Ecology Notes, 7 p: 965-968.

Wang J. Bao A. Gu JHe X. Wu Z. Qiao B. Chen Z. Xiong L. Zhang Y. Zheng H. Li F. Zhao Z. Mei S. Tong Y. 2021. A retrospective analysis of 12,400 SARS-CoV-2 RNA tests in patients with COVID-19 in Wuhan. Medicine (Baltimore) 21, 100(20): e25916.

Xia X. 2018. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution, Mol. Biol. Evol. 35(6) p: 1550–1552.

Xia X 2013. Department of Bio. DAMBE5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. Department of Biology and Center for Advanced Research in Environmental Genomics. University of Ottawa, Ottawa, Ontario, Canada. 9p..

Xia X. Xie Z 2001. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. Journal of Heredity, Vol. 92, Issus 4 p: 371-373.

Xia X. 2000. Data Analysis in Molecular Biology and Evolution. Edi. Springer.

Yoann M 2012. Les arbres phylogénétiques : construction et interprétation. Bioinf-fr.net Geekus biologicus.

Zi-wei Y. Shuofeng Y. Kit-san Y. Sin-yee F. Chi-ping C. Dong-yan J 2020. Zoonotic origins of human coronaviruses. Int. J. Biol. Sci., 16(10) p: 1686-1697.

Abdelguerfi A. Chehat F. Ferrah A. Yahiaoui S 2009. Quatrième rapport national sur la mise en œuvre de la convention sur la diversité biologique au niveau national. Rapport du Ministère de l'aménagement du territoire, de l'environnement et du tourisme algérien.