

UNIVERSITY OF PÉCS

FACULTY OF SCIENCES

INSTITUTE OF BIOLOGY

**SARS-CoV-2 pandemic: phylodynamic, phylogeographic,
and mutation analysis of Algerian sequences**



External Supervisor: Dr.Zeghib Safia

Author: Fatima Ezzohra Ezahedi

Internal Supervisor: Pr.Fekete Csaba

Biology MSc student

Table des matières

1	Introduction:	4
1.1	The human Coronaviruses:	4
1.2	SARS-CoV-2 genome structure:	6
1.3	Phylogenetic studies:	7
1.3.1	Phylodynamic analysis:	8
1.3.2	Phylogeography analysis:	8
2	Aims	10
3	Material and methods:	11
3.1	Choosing sequences and maximum likelihood phylogeny:	11
3.2	Temporal signal assessment:	11
3.3	Time-calibrated phylogeny, Continuous phylogeographic analysis:	11
3.4	Genomic analysis:	12
3.4.1	<i>Synonymous and non-synonymous mutation analysis:</i>	12
3.4.2	Effect of mutation on SARS-CoV-2 proteins:	12
3.5	Haplotype network analysis:	12
4	Results:	13
4.1	Phylogenetic analysis:	13
4.2	Selection pressure:	14
4.3	Mutation Analysis:	17
4.4	Haplotype Network Analysis:	19
5	References:	22

Abstract :

Since the appearance of the SARS-CoV-2 pandemic in 2019 which started in Wuhan China then spread all over the world. More importance is accorded to the genetic and molecular aspects of the disease as it shapes the vaccine and drug developments in addition to the detection of different variants and mutations to effectively implement more control measures to hamper the spread of the disease and evaluate the effect of the mutations on the vaccinations status. However, most of the studies and results are reported from America, Europe and China leaving Africa with a huge knowledge gap. For that purpose, we initiated our work on Algerian samples analyzing a dataset comprising 355 complete genomes and a sub high quality dataset encompassing 193 full complete genomes to cover the genetic and molecular aspect of the pandemic. Several programs were used to implement phylogenetic phylogeographic and phylodynamic studies in addition to mutation assessment and the evolutive selection pressure. Our results demonstrated the high diversity across the Algerian samples, the impact of local and international travels on the spread of the disease in addition to the importance of missing unsampled data in understanding the transmission loop. Moreover, several Algerian characteristic mutations were observed with different effects on the corresponding proteins ranging from neutral to deleterious.

1 Introduction:

Viral zoonoses, is a field of research, where it's main object is to examine the connections between the health of animals, humans, and the environment. This field is important because many diseases, including some of the most dangerous and deadly, are caused by pathogens that can jump from animals to humans. Therefore, in light of the last events the world has faced, the SARS-CoV 2 virus has been getting a lot of attention. There have been many studies that have investigated the evolutionary properties of SARS-CoV-2. However, most of them were analysis of European, Asian, and American countries leaving a gap in the data representing African countries. That's why in our work where we are providing more knowledge about the Algerian pandemic and providing important information about how the virus is changing overtime, while looking deeper into the gene functions and the recombination's that occurred.

1.1 The human Coronaviruses:

Coronaviruses are a diverse group of viruses that can cause a range of symptoms, from mild common cold-like symptoms to severe illnesses like SARS. Some people may experience only mild symptoms, like a runny nose or cough, while others may develop more serious conditions such as fever and pneumonia. These viruses can be particularly dangerous for older individuals or those with underlying health conditions, and in some cases, they can even be fatal (WHO, accessed on 20 January 2023).

Symptomatic transmission occurs when an individual with symptoms spreads the virus through close contact with respiratory droplets, direct contact with the infected person, or contact with contaminated surfaces and objects. Pre-symptomatic transmission happens when an individual spreads the virus before showing symptoms. It has been reported that some people who tested positive for the disease transmitted it 1-3 days before experiencing symptoms (WHO, accessed on 20 January 2023).

Studies have shown that asymptomatic individuals can have similar viral loads as those with symptoms and can spread the virus to others. This has made controlling the spread of the virus

more challenging, as individuals who are unaware, they have the virus, may unknowingly spread it to others (A.Samaha, 2021).

CoVs are classified under the order *Nidovirales*, Family *Coronaviridae*, and Subfamily *Orthocoronavirinae* (Figure 1). For SARS-Cov-2, Next generation sequencing showed 88% homology to two bat-derived CoV-like coronaviruses, 79% homology to SARS-CoV and 50% to MERS-CoV. This phylogenetic analysis showed that SARS-CoV-2 belongs to the Genus of Betacoronaviruses, and the Sarbecovirus Subgenus (Malik, 2020).

The human coronaviruses go back to a long time ago, as the first one was discovered in 1965 from the nasal discharge of a patient with common cold. Mainly, we know six different strains of CoV to infect humans. As shown in the 1st figure HCoV-229E, HCoV-OC43, SARS-CoV, HCoV-NL63, HCoV-HKU1, and Middle East respiratory syndrome coronavirus MERS-CoV. HCoV-229E and HCoV-OC43 are the best characterized as they are the prototype viruses from the two main lineages alpha and beta, and 15%-29% of the time they are cause of common cold (Malik, 2020).

As known, SARS-CoV was behind the outbreak of the severe respiratory disease through China from 2002 to 2003, and MERS-CoV is responsible for the outbreak of the same disease in the Middle east in 2012 (Malik, 2020; Ortego J, 2002).

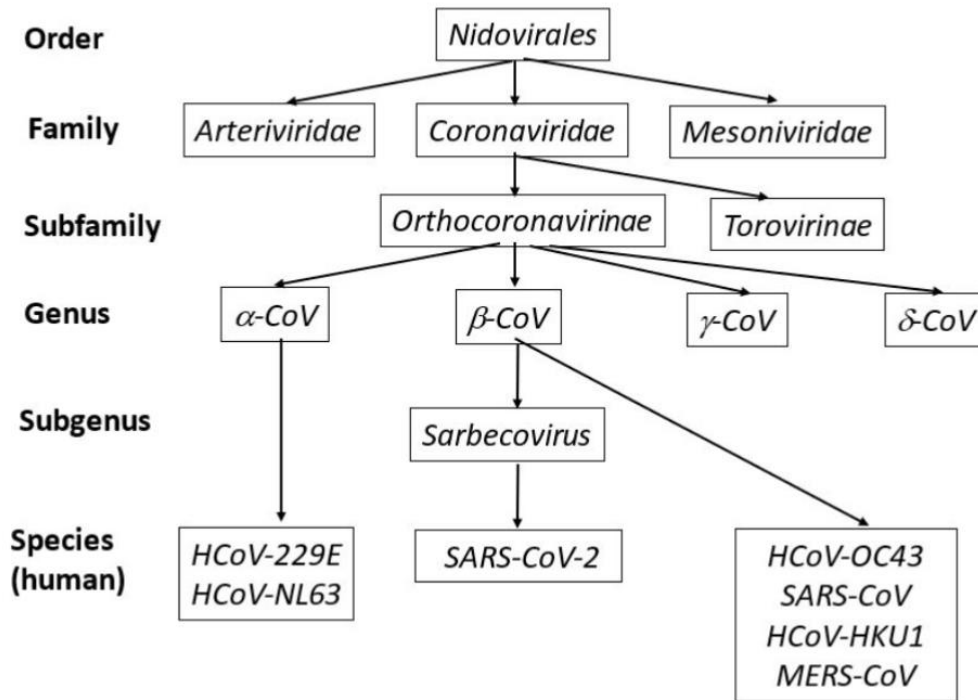


Figure 1: Classification of human Coronaviruses (Malik, 2020)

1.2 SARS-CoV-2 genome structure:

The SARS cov2 genome has approximately 30000 base pairs in length and it encodes for 4 structural proteins, and 21 nonstructural proteins. The genes for these proteins are Orf1a, Orf1b, S gene, Orf3a, E gene, M gene, Orf6, Orf7a, Orf7b, Orf8, N gene, and Orf10 as shown in the 2nd figure. First, the Orf1ab encodes for the NSP1 protein that binds to the 40S subunit of the ribosome and blocks the host mRNA entry channel (Simeoni M, 2021).

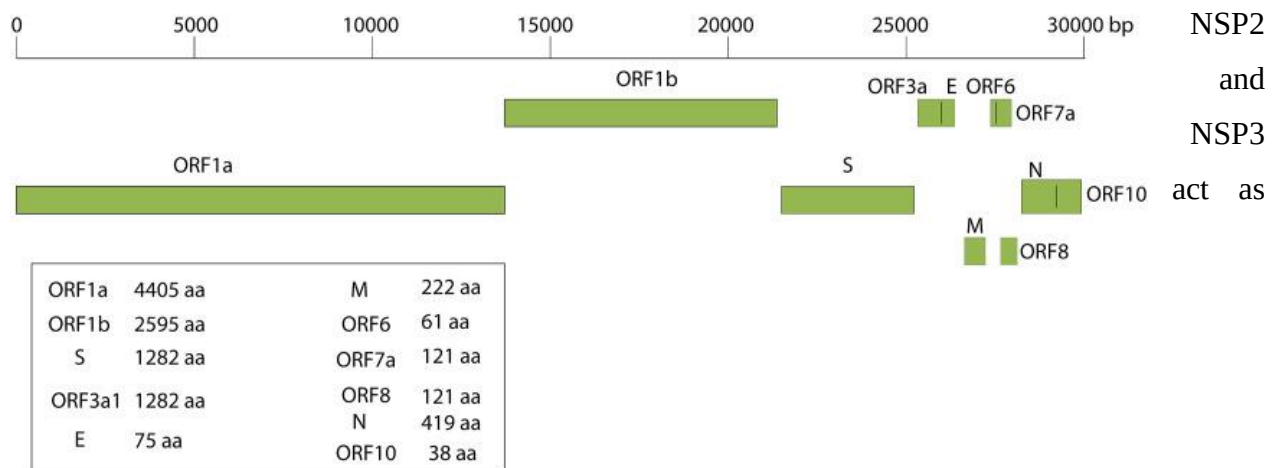


Figure 2: SARS-CoV-2 coding genes (Giovanetti M, 2021)

transmembrane helical segments, and it was prove that the stabilizing mutation falling in the endosome-associated-protein-like domain of the nsp2 protein could account for COVID-2019 high ability of contagious, while the destabilizing mutation in nsp3 proteins could suggest a potential mechanism differentiating COVID-2019 from SARS (Angeletti S, 2020).

NSP4 to facilitate extracellular mtDNA release it induces formation of outer membrane macropore and inner membrane vesicle, alongside with Orf9 (Faizen Mi, 2022). For NSP5, it cleaves off the 10 most-N-terminal amino acids from RIG-I and deprives it of the ability to activate MAVS (mitochondrial antiviral signaling), and potentially inhibits interferon (IFN) induction by double-stranded RNA (dsRNA) in an enzyme-dependent manner (Liu Y, 2021).

NSP6 through oligomerization and an amphipathic helix, zippers ER membranes and establishes the connectors (Ricciardi S, Guarino AM, 2022). NSP7 and NSP8 are accessory factors of RdRP, NSP8 functions like a primase and an RNA 3'-terminal Adenylyltransferase activity. Also, NSP9 is an ssRNA binding protein.

NSP10 which joins NSP16 in a complex, in order to methylate the first nucleotide of a capped RNA strand at the 2'-O position (Benoni R, 2021). And NSP14 in complex with NSP10 provides a proofreading function due to the exonuclease activity of NSP14 (Moeller NH, 2022).

NSP13 was found to inhibit the activity of the IFN-I pathway through forming a complex with TBK1 and p62 (Sui C, 2022). Furthermore, NSP15 is a uridine-specific endoribonuclease help the virus evade triggering an innate immune response (Wilson IM, 2022).

1.3 Phylogenetic studies:

The SARS -CoV-2 virus has a high rate of genetic variation, which has led to the emergence of many different viral variants. These variants, such as Omicron have been shown to have increased transmissibility or resistance to vaccines or therapeutics (He X, 2020).

Phylogenetic analysis has been used to infer the evolutionary relationships between the different strains of SARS-Cov-2 and to track their spread around the world. The information inferred from phylogenies can be used to track the spread of viral outbreaks, identify the origin of new viral strains, and understand how viruses evolve overtime. Additionally, the phylogenetic trees can also be useful in understanding how the virus has evolved to be more transmissible or more pathogenic, and how the spread happens between different hosts (Kuhls K, 2019).

1.3.1 Phylodynamic analysis:

Phylodynamic analysis is the study of evolutionary dynamics of pathogens overtime, it combines phylogenetics and population dynamics that is mainly about how populations change overtime. By analyzing the genetic sequences of different viral strains collected through a population from different individuals and at different times. Phylodynamics can also be used to infer the basic reproductive number (R_0) of a virus, which is the number of secondary cases produced by a single primary case in a fully susceptible population (Shaw CL, 2021).

Moreover, by analyzing the genetic sequences of viral strains collected before and after an intervention this analysis can determine how the virus evolved in response to the interventions. Studies have shown that the SARS-CoV-2 virus has a high R_0 and that interventions like lockdowns and vaccination have affected its dynamics and evolution.

1.3.2 Phylogeography analysis:

Another necessary part of phylogenetic studies is phylogeographies. Phylogeography is the study of geographical distribution of genetic variation within a species. It's a combination of phylogenetics and biogeography which represents the living organism's geographical distribution. Just like its name suggests, this analysis can be used to understand how a virus spreads and evolves overtime and space. This identifies the geographical origin of the virus and tracks its spreading (Emerson BC, 2011).

For Covid-19, the phylogeography of the virus shows that maybe it has originated in bats and then transmitted to humans in a wet market in Wuhan, China (Holmes EC, 2021).

The COVID-19 had a significant impact on Algeria, as it has in many countries around the world. In Algeria, the first confirmed case was on February 25th, 2020. The government quickly implemented measures to contain the spread but despite these measures the number of cases and deaths has risen steadily. In late 2020 and early 2021, the numbers continued to increase, leading to the government re-imposing some restrictions in several regions of the country. And in January 2021, the first doses of vaccines have been administered starting with healthcare workers.

In previous studies, evolutionary analysis was conducted on a data set of 95 genomes compromising Algerian and other sequences from all over the world, showing multiple introduction of the virus from other countries despite the lockdown, also they were able to identify

an individual amino acid replacement such as a deleterious replacement A23T in the Orf3a gene in an Algerian sequence (Zeghib S, 2021).

However, this study contained only 29 Algerian genome sequences so more sequences would definitely give off more information. Hence, to understand and give more insight about the pandemic in Algeria, in our work we used 192 SARS-CoV-2 Algerian

Genomes extracted from the GISAID database, and analyzed them using phylogenic approaches, first we conducted a phylodynamic and phylogeographic study of our data using the BEAST software and its associated softwares. Then, a haplotype network analysis was created using PopArt it will demonstrate the multiple introductions and show local transmissions. Furthermore, various softwares and programs were used for further genome explorations.

2 Aims

First, we would like to trace the evolution of SARS-CoV-2 over the course of 2020 to 2023. Then, we would like to study the mutational heterogeneity at the nucleotide and protein level across viral genomes, specifically SARS-CoV2, to highlight both unique and common mutations in Algeria. Moreover, the investigation of the evolution of the virus and how different it is in certain subpopulations which leads to its different variants. Furthermore, conduction of a haplotype network analysis showing recombination events detected within the population. These will give us more insight about the host-pathogen interactions and how pathogens adapt to host immunity, which can help in the design of new vaccines and treatment strategies.

3 Material and methods:

3.1 Choosing sequences and maximum likelihood phylogeny:

Sequences were extracted from the GISAID database, two types of filters were used complete genomes, this chooses only sequences that have more than 29000 nucleotides, and secondly all low coverage sequences were excluded, we made sure to choose the ones that have no N (undefined nucleotide) chain. Next, the sequence alignment was done using MAFFT webserver with a fast and rough FFT-NS-2 parameter. A maximum likelihood tree was executed using IQTREE software, and the best substitution selected was GTR+F+R2.

3.2 Temporal signal assessment:

The treefile from the IQTREE output was used as an input file for TempEst, to test the temporal signal of our data. TempEst is a tool for investigating clocklikeness of molecular phylogenies, it can analyze trees where all sequences have been collected at the same time, contemporaneous trees, and Dated-tip trees which have sequences that were collected at different dates. We use it to analyze if the assumption that our tree is under a molecular clock is true or not, which means if the the evolution of our sequences in the timeline specified is significant or not. The regression analysis of root-to-tip genetic distances in relation to sampling times showed strong positive correlation coefficients $r=0.96$, also an $R^2=0.92$, this indicates that our data is suitable for a phylogenetic molecular clock analysis. It also provides the tree with a best fitting root that is likely the most compatible with the assumption of the molecular clock.

3.3 Time-calibrated phylogeny, Continuous phylogeographic analysis:

BEAST (Bayesian Evolutionary Analysis Sampling Trees), a software package that uses Bayesian statistical methods to infer the evolutionary history of viruses and particularly useful for analyzing molecular sequence data to infer the evolutionary relationship between different strains of the virus. The core of Beast is a probabilistic framework for inferring the evolutionary history of a group of organisms by sampling from the space of possible evolutionary scenarios., this is done using a technique called Markov chain Monte Carlo (MCMC). We used BEAST v1.10.4 for generating the tip-dated phylogenetic trees for the nucleotide substitution model a GTR+F model was used following Gamma type distribution, and for the continuous trait where we used the latitude and longitude of the Algerian cities, the trait model is a Brownian random walk model. This is done under a lognormal uncorrelated relaxed clock. For the tree prior a coalescent Bayesian skyline, it

calculates under the assumption that the sequences are a small sample from a haploid population evolving under wright-Fisher dynamics. The MCMC chains were run for 50 million generations and sampled every 5000 generations. Afterwards, a software named TRACER v1.6.0 is used for examining the effective sampling size that has to be higher than 200. Then using TreeAnnotator v1.10.4 the maximum clade credibility trees (MCC) were annotated and visualized in FigTree v1.4.4. Subsequently, we used Spread3 v0.9.7 software to visualize the transmission routes considering a continuous space.

3.4 Genomic analysis:

3.4.1 Synonymous and non-synonymous mutation analysis:

To evaluate the selective pressure at the protein level SNAP v2.1.1 software was used, it's a bioinformatic tool that analyzes genetic data to classify genetic variations, mutations, such as single nucleotide polymorphisms (SNPs). By comparing sequences, it identifies if the mutation is non-synonymous, change of the Amino Acid sequence, or synonymous, meaning no change. A non-synonymous mutation is deleterious when the change happens with physiochemically different amino acids. Therefore, the probability for them to become fixed in the population is very low leading to and adverse selection. However, when the non-synonymous substitution is advantageous, they are more likely to be fixed in the population and there is more change in the amino acids of the protein. Using SNAP, we got a ω ratio representing the rate of the non-synonymous mutations to the synonymous mutations, and the selective pressure was calculated for each sequence pair within all SARS-cov-2 genes: *Orf1a*, *Orf1b*, *S*, *Orf3a*, *E*, *M*, *Orf6*, *Orf7a*, *Orf7b*, *Orf8*, *N*, and *Orf10*. Here we used only 193 sequences from our original dataset that have no unknown nucleotides in their sequences to avoid biased results.

3.4.2 Effect of mutation on SARS-CoV-2 proteins:

The Algerian sequences were analyzed using a mutation app implemented in the GISAID database to highlight the mutational pattern and the variations in the amino-acid level called CoV-surver. Afterwards, we applied these mutations on the reference sequence of each protein in PredictSNP to see the effect of the following changes on the function of the protein and disease relation.

3.5 Haplotype network analysis:

A haplotype network analysis is used in evolutionary studies to infer the relationships among haplotypes, which are sets of alleles on a chromosome that are inherited together. The basic principle is that haplotypes are more similar to each other, so they are more likely to have a recent common ancestor. First using the DnaSP v6.12.03 software, a software tool that is used to analyze various aspects of DNA sequence data, it can estimate recombinations, number of segregating sites, and haplotype generation. Then to make the haplotype, POPART was used, a software that will infer evolutionary relationships among haplotypes using a median-joining algorithm, this will create a graphical representation of the evolutionary relationships among the set of haplotypes, making a haplotype network. The analysis was done using default setting/epsilon=0.

4 Results:

4.1 Phylogenetic analysis:

This plot shows our phylogenetic tree, where blue lines show low posterior probabilities and red lines show high posterior probability, the low posterior probabilities are due to missing data.

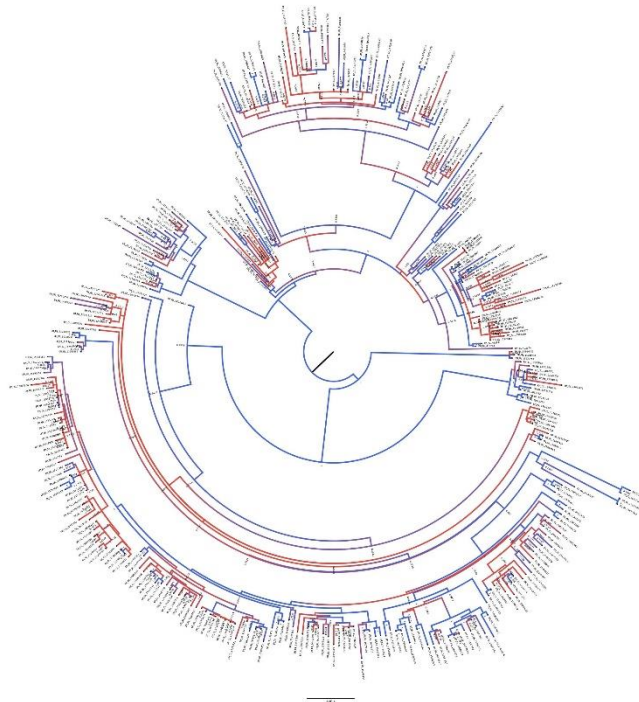


Figure 3: phylogenetic tree of SARS-CoV-2 algerian sequences

4.2 Selection pressure:

To observe the selective pressure on each gene, the data we got from SNAP was gathered in table 1. First, we know that the ω ratio (dn/ds) represents a positive selection if higher than 1, and a negative selection if lower than 1. As shown in the Table 1, we can see that Orf1a with a value of 0,26 is highly negative, alongside with Orf3a, E gene, Orf6, Orf7a, and Orf8. As for Orf1b, M gene, and N gene they showed moderately negative selection. However, the Spike gene had a value of 5.06 which proves highly positive selection. For NSP10 there were no significant results.

Gene	ds	dn	dn/ds
<i>Orf1a</i>	0.0023	0.0006	0.26
<i>Orf1b</i>	0.0013	0.0007	0.53
<i>Spike</i>	0.0016	0.0081	5.06
<i>Orf3a</i>	0.0056	0.0018	0.32
<i>E</i>	0.0179	0.0059	0.32
<i>M</i>	0.0068	0.0053	0.77
<i>Orf6</i>	0.0284	0.0067	0.23
<i>Orf7a</i>	0.0120	0.0036	0.3
<i>Orf8</i>	0.0131	0.0035	0.26
<i>N</i>	0.0058	0.0050	0.86

Table 1: Selection pressure of SARS-CoV-2 genes

ds: observed synonymous substitutions; dn: observed non-synonymous substitutions.

dn/ds: selection pressure.

SNAP also generated plots that show the insertion/ deletion ratio (in/del), synonymous mutations rate (syn), and non-synonymous mutations rates (non-syn). We got results for each gene.

As we can see in figure 3, only Orf1a shows that the in/del ratio is higher than syn and non syn mutations, some of the other genes such as Orf1b, E, M, Spike and Orf8 show the rate of non-syn to be higher. As for Orf3a, Orf7a, and Orf6 they show the syn mutation rates to be much more frequent than non-syn.

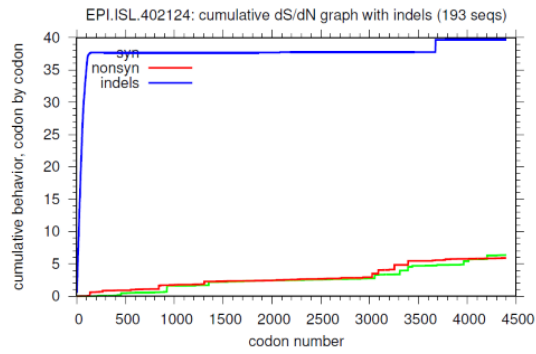
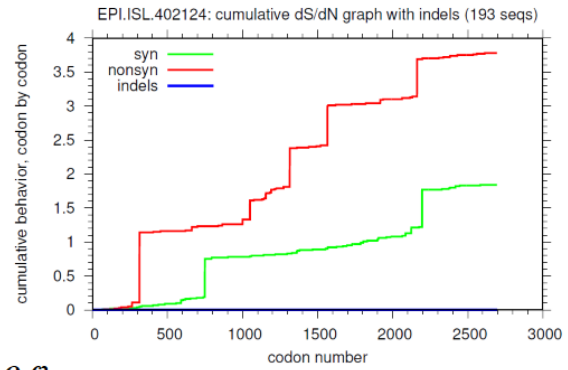
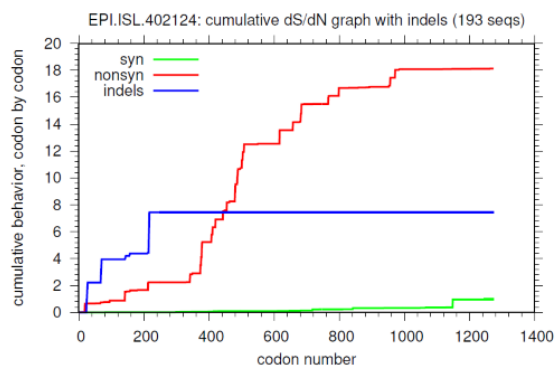
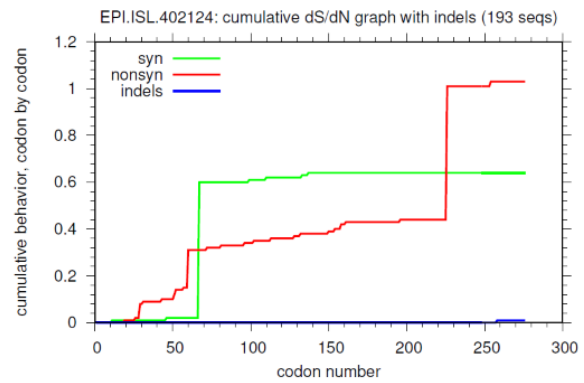
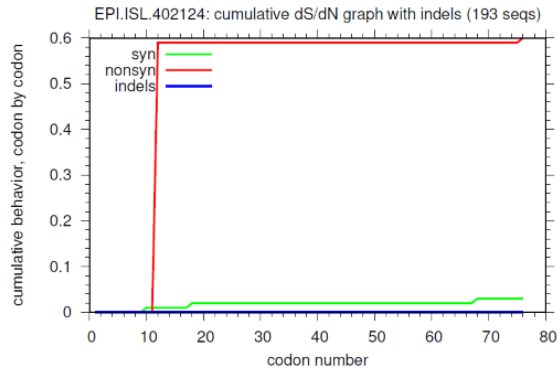
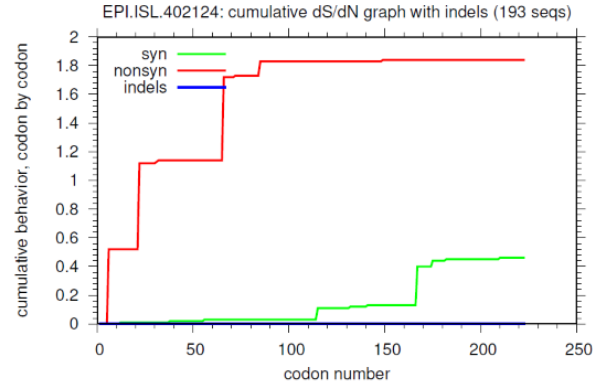
ORF1A**ORF1B****SPIKE:****Orf3a**

Figure 4: Plots for insertion/ deketions, synonymous and non-synonymous rate in each nucleotides in the genes (part1)

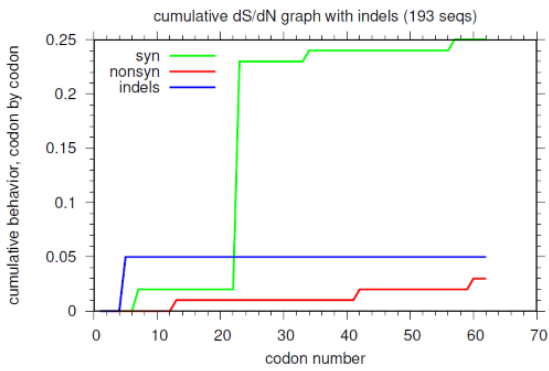
E



M



ORF6



ORF7A

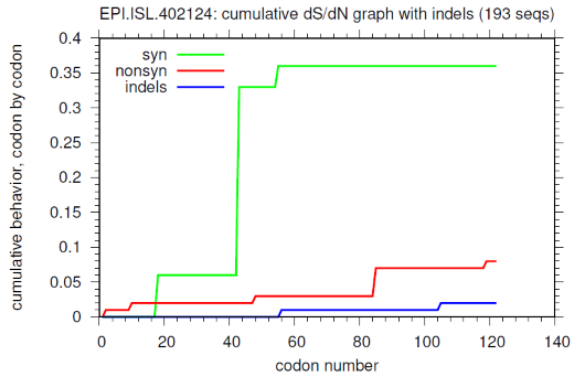
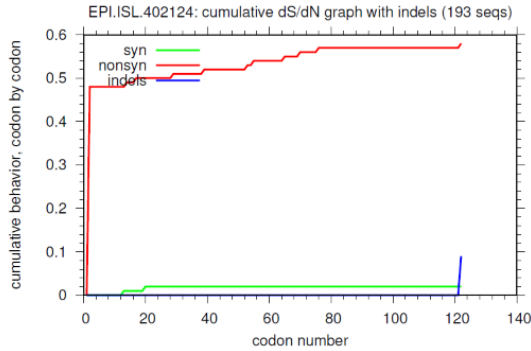
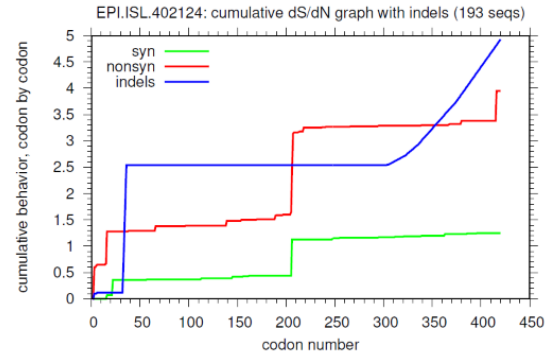


Figure 5: Plots for insertion/ deletions, synonymous and non-synonymous rate in each nucleotides in the genes (part2)

ORF8



NGENE



NSP10

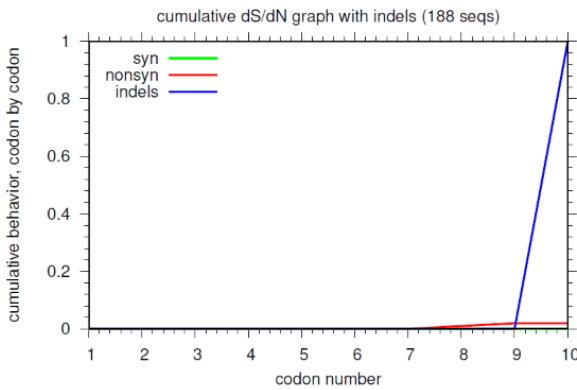


Figure 6: Plots for insertion/ deletions, synonymous and non-synonymous rate in each nucleotides in the genes (part3)

4.3 Mutation Analysis:

Using PredictSNP, we inserted mutations present in Algerian sequences from two timelines, ones from the beginning of the pandemic, and the others that appeared recently in 2022. PredictSNP, gathers information from all other methods shown (Table2), to give a percentage of how deleterious (red color) or neutral (green color) the mutations are. For NSP1, the mutation has been present since the beginning of the pandemic and we can see that it's a deleterious mutation. In NSP2 the most common mutation between our sequences is Q376K and it's a neutral mutation, adding that it's only present in 2022 sequences. Then, NSP3 and NSP6, all their mutations are recent ones, and each of them has one deleterious mutation and the others are neutral. As for, NSP4, NSP5, NSP13, NSP14, NSP15, E, and NSP7b they all have neutral mutations that appear in a large amount of recent sequences. Spike showed neutral mutations for both types of sequences,

old and new ones. And finally, NS3, M, NS8, M seem to have a deleterious mutation each that also appear only in recent sequences, also some neutral mutations (Table 3).

Potein	Mutation	PredictSNP	MAPP	Phd-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
NSP1	S135R	61 %	92 %	66 %	67 %	43 %	53 %	62 %
NSP2	Q376K	60 %	82 %	83 %	67 %	72 %	46 %	56 %
NSP3	T24I	60 %	64 %	72 %	67 %	43 %	45 %	62 %
	T183I	74 %	72 %	83 %	67 %	72 %	45 %	55 %
	G489S	61 %	85 %	55 %	59 %	81 %	79 %	56 %
NSP4	L264F	83 %	65 %	98 %	67 %	68 %	77 %	71 %
	T327I	83 %	85 %	72 %	67 %	68 %	77 %	67 %
	T492I	83 %	76 %	72 %	67 %	70 %	71 %	50 %
NSP5	P132H	75 %	77 %	83 %	67 %	68 %	67 %	58 %
NSP6	Q208R	83 %	63 %	58 %	67 %	68 %	75 %	67 %
	R233C	72 %	48 %	77 %	74 %	68 %	79 %	81 %
NSP13	T127N	83 %	79 %	55 %	67 %	79 %	67 %	77 %
	M233I	83 %	70 %	78 %	67 %	76 %	68 %	55 %
	R392C	63 %	70 %	66 %	67 %	68 %	53 %	62 %
NSP14	I42V	83 %	70 %	72 %	67 %	-	66 %	71 %
NSP15	T112I	83 %	89 %	67 %	73 %	76 %	76 %	58 %

Table 2:Mutations in SARS-CoV-2 proteins part1

Potein	Mutation	PredictSNP	MAPP	Phd-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
SPIKE	A27S	83 %	74 %	83 %	67 %	64 %	74 %	77 %
	N679K	83 %	-	72 %	67 %	74 %	78 %	50 %
	V213G	63 %	64 %	68 %	67 %	63 %	79 %	56 %
	N658S	83 %	78 %	72 %	67 %	64 %	89 %	61 %
	D796Y	75 %	64 %	68 %	67 %	69 %	84 %	58 %
	Q954H	74 %	75 %	78 %	67 %	63 %	65 %	56 %
NS3	S26L	63 %	75 %	83 %	67 %	73 %	79 %	50 %
	T223I	87 %	92 %	68 %	59 %	-	79 %	85 %
E	T9I	75 %	86 %	51 %	67 %	71 %	81 %	61 %
M	D3N	75 %	79 %	98 %	67 %	40 %	70 %	67 %
	Q19E	83 %	78 %	83 %	67 %	72 %	87 %	67 %
	A63T	60 %	46 %	68 %	67 %	55 %	67 %	55 %
	I82T	72 %	77 %	58 %	67 %	63 %	79 %	81 %
NS7b	T40I	83 %		66 %		-		-
NS8	A51V	65 %	63 %	78 %	67 %	71 %	43 %	62 %
	R52I	72 %	84 %	72 %	74 %	60 %	79 %	87 %
N	A12G	68 %	64 %	72 %	67 %	50 %	46 %	77 %
	R203K	63 %	79 %	78 %	67 %	60 %	53 %	72 %
	R203M	72 %	43 %	78 %	74 %	81 %	79 %	85 %
	G204P	72 %	66 %	72 %	74 %	81 %	79 %	56 %

Table 3: Mutations in SARS-CoV-2 proteins part1

4.4 Haplotype Network Analysis:

For the haplotype we got 179 haplotypes, that shows a huge diversity in our dataset, we can see colored nodes which represent Algerian cities we got our samples from, and we got black inter-nodes, these show missing sampling data. Moreover, we can see that data from same locations cluster together, but also data from different cities make clusters together showing the absence of restrictions.

also that negative selection on the other genes can tell us that the function of those genes is very important and needs improvement for the virus to become stronger and spread more.

We can't deny that the amount of samples is very important for viral evolutionary studies, with our data set that has 344 sequences we were able to get this kind of information but still a lot of missing data was needed in this study for more precise results. Previous studies had approximately only 18 fully sequenced genomes, while ours had 193 full sequences and with this data we had a lot more information on the Algerian lineages and we were able to get 179 types of haplotypes in only one country (Zeghib S, 2021).

The clusters shown in the haplotype network showed 2 types, ones from the cities and ones from totally different places this can only show how when restrictions were removed from the country the virus has been roaming around from all different kind of individuals which improved even its mutation rate more and made it more diverse.

Our protein sequences, most mutations were found to be neutral but some of them were deleterious. The R233C mutation found in NSP6 was discovered to be a missense mutation and it decreases the structural stability of the spike protein, and this might decrease the infectivity of the virus (Hossein MS, 2021)

5 References:

Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol.* 2020 Jun;92(6):584-588. doi: 10.1002/jmv.25719. Epub 2020 Feb 28. PMID: 32083328; PMCID: PMC7228367.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J., Damborsky, J., 2014: PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLOS Computational Biology* 10: e1003440.

Benoni R, Krafcikova P, Baranowski MR, Kowalska J, Boura E, Cahová H. Substrate Specificity of SARS-CoV-2 Nsp10-Nsp16 Methyltransferase. *Viruses.* 2021 Aug 30;13(9):1722. doi: 10.3390/v13091722. PMID: 34578302; PMCID: PMC8472550.

Emerson BC, Cicconardi F, Fanciulli PP, Shaw PJ. Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philos Trans R Soc Lond B Biol Sci.* 2011 Aug 27;366(1576):2391-402. doi: 10.1098/rstb.2011.0057. PMID: 21768154; PMCID: PMC3130430.

Faizan MI, Chaudhuri R, Sagar S, Albogami S, Chaudhary N, Azmi I, Akhtar A, Ali SM, Kumar R, Iqbal J, Joshi MC, Kharya G, Seth P, Roy SS, Ahmad T. NSP4 and ORF9b of SARS-CoV-2 Induce Pro-Inflammatory Mitochondrial DNA Release in Inner Membrane-Derived Vesicles. *Cells.* 2022 Sep 23;11(19):2969. doi: 10.3390/cells11192969. PMID: 36230930; PMCID: PMC9561960.

Giovanetti M, Benedetti F, Campisi G, Ciccozzi A, Fabris S, Ceccarelli G, Tambone V, Caruso A, Angeletti S, Zella D, Ciccozzi M. Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochem Biophys Res Commun.* 2021 Jan 29; 538:88-91. doi: 10.1016/j.bbrc.2020.10.102. Epub 2020 Nov 6. PMID: 33199021; PMCID: PMC7836704.

He X, Hong W, Pan X, Lu G, Wei X. SARS-CoV-2 Omicron variant: Characteristics and prevention. *MedComm (2020).* 2021 Dec 16;2(4):838-845. doi: 10.1002/mco2.110. PMID: 34957469; PMCID: PMC8693031.

Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF, Doherty PC, Farrar J, Geoghegan JL, Jiang X, Leibowitz JL, Neil SJD, Skern T, Weiss SR, Worobey M, Andersen KG, Garry RF, Rambaut A. The origins of SARS-CoV-2: A critical review. *Cell*. 2021 Sep 16;184(19):4848-4856. doi: 10.1016/j.cell.2021.08.017. Epub 2021 Aug 19. PMID: 34480864; PMCID: PMC8373617.

Hossain MS, Pathan AQMSU, Islam MN, Tonmoy MIQ, Rakib MI, Munim MA, Saha O, Fariha A, Reza HA, Roy M, Bahadur NM, Rahaman MM. Genome-wide identification and prediction of SARS-CoV-2 mutations show an abundance of variants: Integrated study of bioinformatics and deep neural learning. *Inform Med Unlocked*. 2021;27:100798. doi: 10.1016/j.imu.2021.100798. Epub 2021 Nov 18. PMID: 34812411; PMCID: PMC8598266.

<https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>

Kuhls K, Mauricio I. Phylogenetic Studies. *Methods Mol Biol*. 2019; 1971:9-68. doi: 10.1007/978-1-4939-9210-2_2. PMID: 30980297.

Liu J, Liao X, Qian S, et al. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020. *Emerg Infect Dis*. 2020; 26(6). doi.org/10.3201/eid2606.200239

Liu Y, Qin C, Rao Y, Ngo C, Feng JJ, Zhao J, Zhang S, Wang TY, Carriere J, Savas AC, Zarinfar M, Rice S, Yang H, Yuan W, Camarero JA, Yu J, Chen XS, Zhang C, Feng P. SARS-CoV-2 Nsp5 Demonstrates Two Distinct Mechanisms Targeting RIG-I and MAVS To Evade the Innate Immune Response. *mBio*. 2021 Oct 26;12(5):e0233521. doi: 10.1128/mBio.02335-21. Epub 2021 Sep 21. PMID: 34544279; PMCID: PMC8546575.

Malaysian J Pathol 2020; 42(1) : 3 – 11

Moeller NH, Passow KT, Harki DA, Aihara H. SARS-CoV-2 nsp14 Exoribonuclease Removes the Natural Antiviral 3'-Deoxy-3',4'-didehydro-cytidine Nucleotide from

RNA. *Viruses*. 2022 Aug 16;14(8):1790. doi: 10.3390/v14081790. PMID: 36016411; PMCID: PMC9415739.

Ortego J, Escors D, Laude H, Enjuanes L. Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. *J Virol*. 2002 Nov;76(22):11518-29. doi: 10.1128/jvi.76.22.11518-11529.2002. PMID: 12388713; PMCID: PMC136772.

Ricciardi S, Guarino AM, Giaquinto L, Polishchuk EV, Santoro M, Di Tullio G, Wilson C, Panariello F, Soares VC, Dias SSG, Santos JC, Souza TML, Fusco G, Viscardi M, Brandi S, Bozza PT, Polishchuk RS, Venditti R, De Matteis MA. The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature*. 2022 Jun;606(7915):761-768. doi: 10.1038/s41586-022-04835-6. Epub 2022 May 12. PMID: 35551511; PMCID: PMC7612910.

Samaha AA, Mouawia H, Fawaz M, Hassan H, Salami A, Bazzal AA, Saab HB, Al-Wakeel M, Alsaabi A, Chouman M, Moussawi MA, Ayoub H, Raad A, Hajjeh O, Eid AH, Raad H. Effects of a Single Dose of Ivermectin on Viral and Clinical Outcomes in Asymptomatic SARS-CoV-2 Infected Subjects: A Pilot Clinical Trial in Lebanon. *Viruses*. 2021 May 26;13(6):989. doi: 10.3390/v13060989. Retraction in: *Viruses*. 2021 Oct 26;13(11): PMID: 34073401; PMCID: PMC8226630.

Shaw CL, Kennedy DA. What the reproductive number R_0 can and cannot tell us about COVID-19 dynamics. *Theor Popul Biol*. 2021 Feb;137:2-9. doi: 10.1016/j.tpb.2020.12.003. Epub 2021 Jan 5. PMID: 33417839; PMCID: PMC7785280.

Simeoni M, Cavinato T, Rodriguez D, Gatfield D. I(nsp1)ecting SARS-CoV-2-ribosome interactions. *Commun Biol*. 2021 Jun 10;4(1):715. doi: 10.1038/s42003-021-02265-0. PMID: 34112887; PMCID: PMC8192748.

Sui C, Xiao T, Zhang S, Zeng H, Zheng Y, Liu B, Xu G, Gao C, Zhang Z. SARS-CoV-2 NSP13 Inhibits Type I IFN Production by Degradation of TBK1 via p62-Dependent Selective Autophagy.

J Immunol. 2022 Feb 1;208(3):753-761. doi: 10.4049/jimmunol.2100684. Epub 2022 Jan 7. PMID: 34996837.

Wilson IM, Frazier MN, Li JL, Randall TA, Stanley RE. Biochemical Characterization of Emerging SARS-CoV-2 Nsp15 Endoribonuclease Variants. J Mol Biol. 2022 Oct 30;434(20):167796. doi: 10.1016/j.jmb.2022.167796. Epub 2022 Aug 19. PMID: 35995266; PMCID: PMC9389836.

Zeghib S, Somogyi BA, Zana B, Kemenesi G, Herczeg R, Derrar F, Jakab F. The Algerian Chapter of SARS-CoV-2 Pandemic: An Evolutionary, Genetic, and Epidemiological Prospect. Viruses. 2021 Aug 2;13(8):1525. doi: 10.3390/v13081525. PMID: 34452390; PMCID: PMC8402747.