JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

ARTICLE | DOI: 10.5584/jiomics.v14i1.228

Navigating the molecular diversity of SARS-CoV-2: early pandemic insights from comparative phylogenetic analysis.

Chahinez Triqui[1], Ilyes Zatla[2], Wafaa Lemerini[3], Nora Benmadani[4], Mohammed Charaf Eddine Houari[4], Semir Bechir Suheil Gaouar[4], Nassima Mokhtari-Soulimane[1]

[1]Laboratory of Physiology, Physiopathology and Biochemistry of Nutrition, Department of Biology, Faculty of Natural and Life Sciences, Earth and Universe, University of Tlemcen, Algeria. [2]Laboratory of Microbiology applied to the Food industry, Biomedical and the Environment, Faculty of Natural and Life Sciences, Earth and Universe Sciences. Department of Biology. University of Tlemcen, Algeria. [3]Laboratory of Organic Chemistry, Natural Substances, and Analysis, Faculty of Natural and Life Sciences, Earth and Universe Sciences. Department of Biology. University of Tlemcen, Algeria. [4]Laboratory of Applied genetic in agriculture, ecology and public health, Faculty of Natural and Life Sciences, Earth and Universe, University of Tlemcen, Algeria.

Available Online: 15 April 2024

Abstract

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019 precipitated the onset of the COVID-19 pandemic, which swiftly spread across more than 214 countries and territories, posing a significant global health crisis. In response, laboratories worldwide have embarked on extensive efforts to characterize the genomic landscape of the virus, employing a myriad of sophisticated genomic analysis techniques. This study endeavors to undertake a comprehensive exploration into the genetic diversity, geographical distribution, and virulence determinants of SARS-CoV-2 clades across 11 diverse countries, employing advanced computational biology methodologies. Leveraging molecular data sourced from prominent international databases, the analysis aims to unravel the intricate phylogenetic relationships and mutational dynamics exhibited by various viral strains circulating worldwide. The findings of this investigation promise to yield invaluable insights into the evolutionary trajectory of SARS-CoV-2, shedding light on potential therapeutic targets and informing strategies for mitigating the impact of the ongoing pandemic on global public health. Results highlight significant genetic diversity among SARS-CoV-2 strains across different countries, with phylogenetic analysis revealing distinct subclass groupings within each country. A manual comparison of sequences identified numerous mutations, with certain mutations associated with increased virulence. Comparison of clade G and clade O sequences revealed differences in mutation profiles, suggesting potential links to virulence and transmissibility. These findings underscore the dynamic nature of SARS-CoV-2 evolution and the importance of monitoring genetic changes for public health interventions.

Keywords: SARS-CoV-2, COVID-19, Mutation, Variant, Phylogeny.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel member of the Coronaviridae family, emerged as a global health threat in December 2019, originating from the bustling city of Wuhan, China. Its rapid transmission and virulence swiftly propelled the onset of the COVID-19 pandemic, spreading relentlessly across continents and precipitating unprecedented challenges to public health systems worldwide [1-4]. Notably, SARS -CoV-2 shares genetic similarities with other members of the coronavirus family, including Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) [5-8]. Despite this genetic kinship, SARS-CoV-2 exhibits a distinctive clinical profile, encompassing a broad spectrum of manifestations ranging from mild respiratory symptoms to severe pneumonia and acute respiratory distress syndrome (ARDS). The multifaceted clinical presentation of COVID-19 underscores the imperative for a nuanced understanding of the genetic determinants underlying its pathogenesis and transmission dynamics [9-12].

Central to unraveling the complexities of SARS-CoV-2 infection is an elucidation of its genetic diversity and geographical distribution. The genetic variability exhibited by different SARS-CoV-2 strains holds pivotal implications for tracking transmission patterns, elucidating disease dynamics, and identifying potential therapeutic

***Corresponding author:** Ilyes Zatla, ilyes.zatla@univ-tlemcen.dz, Tel: +213-5403-15422

targets [13]. Against this backdrop, this study endeavors to undertake a comprehensive analysis of the genomic sequences of SARS-CoV-2 strains sourced from diverse geographical locations. By harnessing advanced computational biology methodologies, the investigation seeks to delineate the evolutionary dynamics and virulence determinants inherent to distinct viral lineages.

Specifically, our research provides valuable insights into the genetic diversity and mutational dynamics of SARS-CoV-2 strains, which are critical factors influencing the virus's antigenic properties and its ability to evade host immune responses. By characterizing the genomic evolution of the virus and identifying potential virulence determinants, our study contributes to the broader understanding of how SARS-CoV-2 interacts with the immune system and how these interactions may impact disease severity, transmission dynamics, and vaccine efficacy. Moreover, the identification of distinct viral clades and mutation profiles across different geographic regions underscores the importance of ongoing surveillance and molecular epidemiology efforts, which are essential for guiding immunization strategies and vaccine design initiatives. By elucidating the evolutionary trajectory of SARS-CoV-2, our research provides valuable insights that are pertinent to the field of immunology and have implications for public health interventions aimed at controlling the COVID-19 pandemic.

Materials and methods

Genome sequences of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) originating from 11 countries (Algeria, Germany, Australia, England, Spain, France, Italy, Saudi Arabia, Kuwait, Switzerland, USA) were retrieved from the Global Initiative on Sharing All Influenza Data (GISAID) database. A comprehensive dataset comprising a total of 298 SARS-CoV-2 sequences was assembled for subsequent analysis.

Sequence Analysis: Bioinformatics tools were employed for comprehensive sequence analysis, encompassing sequence alignment, phylogenetic tree construction, and mutation analysis. From those softawares we manged to work with MEGA 7, a desktop application for molecular evolutionary genetics analysis, facilitates the analysis of homologous gene sequences from multigene families or different species, focusing on inferring evolutionary relationships and DNA/protein evolution models [14]. Gblocks, a computer program, selectively removes poorly aligned positions and divergent regions from DNA or protein sequence alignments, enhancing the alignment quality for subsequent phylogenetic analysis. It follows reproducible conditions to select blocks based on conservation and gap density criteria, facilitating automation and reproducibility of phylogenetic analyses [15]. Additionally, GENIEGEN software allows for the analysis of DNA, RNA, and protein sequences, aiding in the discovery of genetic information expression, genotype-phenotype relationships, gene polymorphism, multigene families, and predictions in human genetics. It functions as a database of nucleic and peptide sequences, with the capability to incorporate new sequences [16]. Sequences were aligned using state-of-the-art alignment algorithms to ensure accurate alignment across the dataset. The resulting alignment served as the foundation for subsequent phylogenetic analyses.

Phylogenetic Analysis: Phylogenetic trees were constructed using robust methodologies to elucidate the evolutionary relationships among SARS-CoV-2 strains. Phylogenetic tree construction involved iterative processes, with sequences grouped based on the geographical location and data size of the countries under study to facilitate comparative analysis: Group 1: USA; Group 2: UK (United Kingdom); Group 3: AUKA (Australia, Kuwait, Saudi Arabia, Algeria); Group 4: GISA (Germany, Italy, Switzerland); Group 5: FESP (France, Spain). The construction of phylogenetic trees aimed to delineate the evolutionary dynamics of SARS-CoV-2 strains, providing insights into their geographical distribution and evolutionary origins.

Mutation Analysis: Mutations within SARS-CoV-2 genomes were systematically identified and analyzed to assess their potential impact on viral virulence and transmission dynamics. Comparative analysis of mutations between different clades enabled the identification of key genetic determinants associated with disease severity and transmissibility. Mutational landscape analysis provided critical insights into the evolutionary trajectory of SARS-CoV-2 and its adaptive mechanisms in response to selective pressures.

Statistical Analysis: Statistical methodologies were employed to quantify the significance of observed mutations and to assess their potential association with clinical outcomes. Comparative analyses between different clades and geographical regions were conducted to identify statistically significant differences in mutation frequencies and distributions.

Ethical Considerations: This study adhered to ethical guidelines for the use of genomic data, ensuring compliance with data-sharing policies and privacy regulations. All genomic data were anonymized and obtained from publicly available databases, with no identifiable information included in the analysis.

Results and Discussion

In our study, we focused on the phylogenetic analysis and comparison of COVID-19, which has garnered significant media coverage since its emergence in December 2019. Numerous laboratories have dedicated considerable time and effort to characterizing the virus using multiple genome-based techniques, aiming for a comprehensive understanding of the SARS-CoV-2 genome [17-20]. Nearly 300 sequences from various countries across different continents, obtained from the public GISAID database [21], were analyzed to achieve a clear resolution of the virus's diversity, evolution, mutations, and their positions within its genome.

Phylogenetic Analysis

Phylogenetic analysis of SARS-CoV-2 genomes unveiled a landscape rich in genetic diversity, reflecting the complex evolutionary dynamics of the virus. Within each country, distinct subclass groupings were discerned, underscoring the diverse evolutionary trajectories of SARS-CoV-2 strains across different

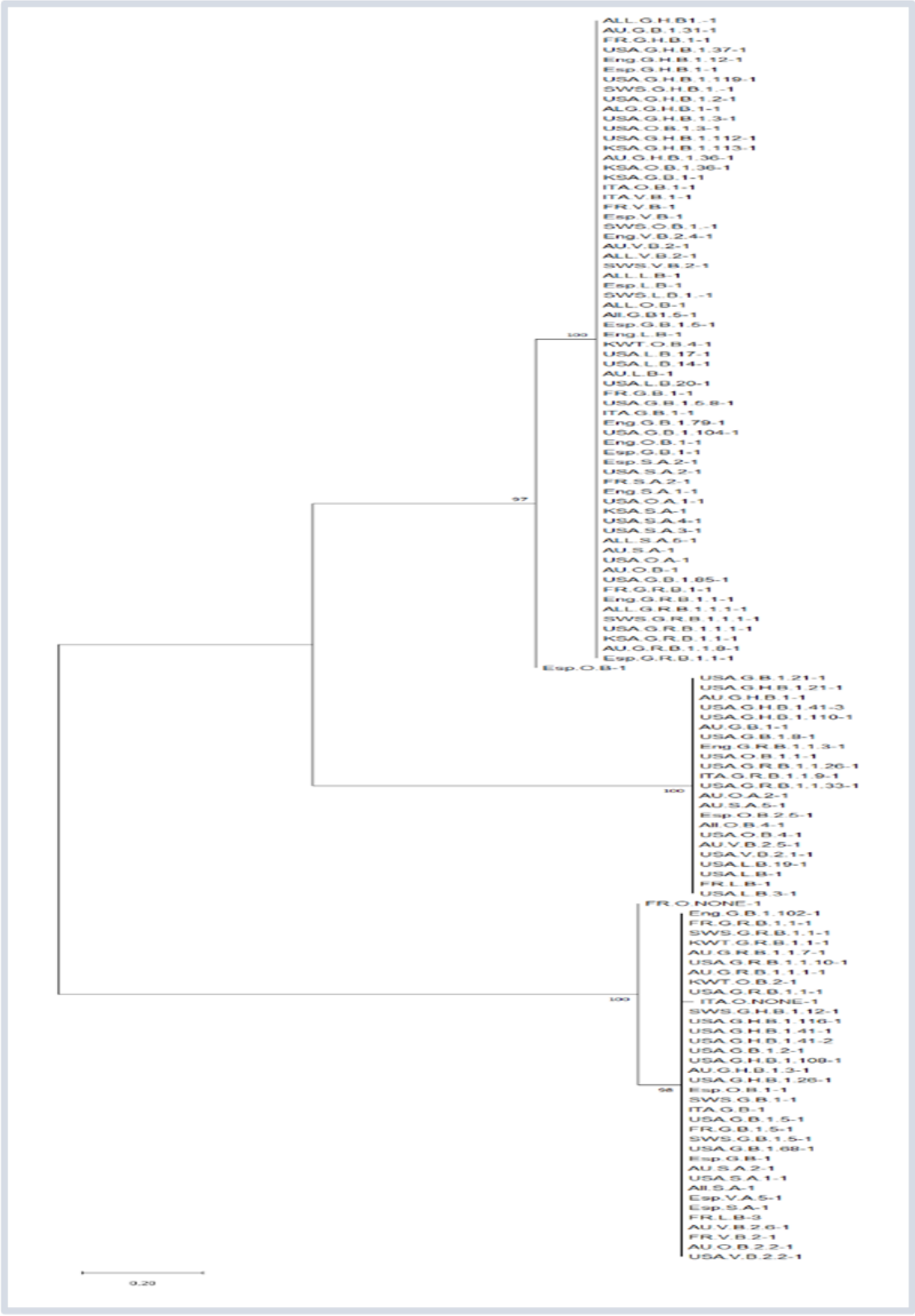

Figure 1: The phylogenetic tree grouping all countries of this study.

geographical regions. This intricate phylogenetic architecture highlights the dynamic nature of viral evolution and the potential for localized viral adaptation in response to environmental and host factors. Delving deeper into the phylogenetic trees, distinct patterns emerge within each country or region studied. For instance, in the USA (Group 1), a diverse array of viral lineages, encompassing clades G, GH, GR, S, L, V, and O, was observed, reflecting the complex epidemiological landscape of the pandemic. Similarly, phylogenetic analyses for other groups (ANG, AU, FE, ALL) unveiled a mosaic of viral lineages, underscoring the genetic heterogeneity inherent to SARS-CoV-2. The final phylogenetic tree (Figure 1) encompasses the results of previous steps and all the countries studied. It is divided into two classes, each further divided into subclasses grouping the different lineage sequences from the countries. We observed in the phylogenetic trees of this study that at the level of each subclass, there is no clear association with the region, lineage, or clade, reflecting the immense diversity and high mutability of this virus.

Mutation Analysis

A manual comparison of SARS-CoV-2 sequences using GenieGen software revealed a myriad of mutations scattered across the viral genome, indicative of ongoing genetic diversification. Notably, several mutations were identified that exhibited a significant association with increased virulence, implicating them as potential determinants of disease severity. Of particular interest were mutations observed in clade G and clade O sequences, which displayed distinct mutation profiles suggestive of differential virulence and transmissibility. These findings underscore the complex interplay between viral genetic variation and disease pathogenesis, highlighting the need for continued surveillance and monitoring of SARS-CoV-2 mutations for effective public health interventions. Table 1 presents the mutations identified through manual comparison of sequences from clades G and O for Group I countries, where mutations are observed in both clade O and clade G, as well as some mutations common to both clades, affecting various key viral proteins, such as ORF1ab, Spike (S) protein, and nucleocapsid (N) protein, suggests their potential role in shaping viral fitness and host interactions.

Table 2 shows the mutations determined from the manual comparison of sequences from clades G and O for Group II countries, where some mutations are exclusive to clade O, others to clade G, and some common to both. Additionally, there are a few mutations shared between Groups 1 and 2. Among the comparison results, 283 mutation positions are identified in the two preceding tables, along with 926 positions of rare mutations, indicating the high mutation rate this virus can undergo. Based on these findings, we suggest that mutations present in clade O may lead to a decrease in virus virulence. The differences in results between the two groups of countries allow us to conclude the significant diversity of this virus. Furthermore, comparative analyses between clades G and O shed light on specific mutations associated with each clade, hinting at potential differences in virulence and transmissibility. Noteworthy mutations identified in Group I and Group II countries, spanning crucial viral proteins, offer tantalizing insights into the evolutionary forces driving the diversification of SARS-CoV-2.

Implications for Public Health

The dynamic nature of SARS-CoV-2 evolution underscores the importance of vigilant surveillance and monitoring of genetic changes for the development of targeted public health interventions. By elucidating the genetic determinants of viral virulence and transmissibility, we can better inform the design of therapeutic strategies and vaccine development efforts. Moreover, the identification of specific mutations associated with increased mortality rates provides valuable insights into potential targets for therapeutic intervention, paving the way for the development of precision medicine approaches tailored to individual patient needs. Importantly, our findings corroborate previous studies [22, 23] suggesting differential virulence between clades G and O. The identification of specific mutations linked to virulence underscores the urgent need for targeted therapeutic interventions. Molecular docking studies targeting key mutations, particularly those associated with heightened virulence, hold promise as a strategy to mitigate the impact of the COVID-19 pandemic.

The study has several limitations that warrant consideration. Firstly, reliance on genomic data sourced from public databases introduces the potential for sampling bias, as certain geographic regions or demographic groups may be overrepresented or underrepresented. Additionally, variability in the accuracy and completeness of genomic data, along with potential sequencing errors or artifacts, could impact the reliability of mutation calls and phylogenetic reconstructions. While the study provides a snapshot of SARS-CoV-2 evolution at a specific time, ongoing viral evolution may lead to changes in genetic diversity and evolutionary relationships over time. Furthermore, establishing causal relationships between genetic variation and clinical outcomes requires additional experimental validation and clinical correlation studies, highlighting the need for caution in interpreting associations between mutations and virulence. Methodological constraints, such as algorithmic biases or assumptions, may also affect the robustness of the findings, necessitating careful validation and sensitivity analyses. Moreover, the generalizability of the findings may be limited to the specific dataset and analytical methods used, requiring replication in independent datasets and diverse populations for validation. Ethical considerations regarding data privacy, informed consent, and data sharing must also be addressed to safeguard individual rights and privacy. Finally, interpretation of phylogenetic trees and mutation profiles may be subject to bias, highlighting the importance of transparency and rigor in reporting methodologies and results.

Conclusions

In summary, this study offers a comprehensive examination of the genetic landscape, geographical distribution, and virulence characteristics of SARS-CoV-2 strains across a diverse array of countries. The intricate phylogenetic patterns observed underscore the dynamic nature of viral evolution and the capacity for adaptation to various environmental pressures. The results of this study

Table 1: Different mutations to identify from the comparison of the sequences of the countries of the Group 1.

Position	Clade G	Clade O	Mutation type	Repetition number	Region
8788	+	+	Substitution T/C	8 (O) / 2 (G)	ORF1ab
14811	-	+	Substitution T/C	10	ORF1ab
11089	-	+	Substitution T/G	9	ORF1ab
14414	+	+	Substitution C/T	10 (O) / 3 (G)	ORF1ab
20274	+	+	Substitution G/A	3 (O) / 9 (G)	ORF1ab
23394	-	+	Substitution A/G	4	Protein S
23403	+	+	Substitution G/A	15 (O) / 4 (G)	Protein S
23409	+	+	Substitution A/G	11 (O) / 1 (G)	Protein S
24868	+	-	Substitution G/A	4	Protein S
26150	-	+	Substitution T/G	8	ORF3a
25435	+	-	Substitution T/G	4	ORF3a
25569	+	+	Substitution T/G	4 (O) / 3 (G)	ORF3a
26720	-	+	Substitution C/G	2	Protein M
26536	+	-	Substitution C/A	2	Protein M
28083	-	+	Substitution C/G	2	ORF8
28150	-	+	Substitution C/T	6	ORF8
28151	-	+	Substitution C/T	2	ORF8
28317	-	+	Substitution C/T	5	Protein N
28694	-	+	Substitution C/T	3	Protein N
28688	-	+	Substitution C/T	3	Protein N
28881	+	+	Substitution A/G	6 (O) / 2 (G)	Protein N
28882	+	+	Substitution A/G	5 (O) / 2 (G)	Protein N
28883	+	+	Substitution C/G	5 (O) / 2 (G)	Protein N
28887	+	+	Substitution A/G	1 (O) / 3 (G)	Protein N
28888	+	+	Substitution A/G	1 (O) / 3 (G)	Protein N
28889	+	+	Substitution C/G	1 (O) / 3 (G)	Protein N

Table 2: The different mutations to be identified from the comparison of the sequences of the countries of Group 2.

Position	Clade G	Clade O	Mutation type	Repetition number	Region
3028	-	+	Substitution C/T	6	ORF1ab
3031	-	+	Substitution T/C	7	ORF1ab
3037	+	+	Substitution C/T	17 (O) / 4 (G)	ORF1ab
3336	-	+	Substitution C/T	8	ORF1ab
11083	-	+	Substitution T/G	11	ORF1ab
14805	-	+	Substitution T/C	4	ORF1ab
14408	+	+	Substitution C/T	10 (O) / 6 (G)	ORF1ab
23380	-	+	Substitution A/G	4	Protein S
23394	-	+	Substitution A/G	4	Protein S
23403	+	-	Substitution G/A	16 (O) / 4 (G)	Protein S
26720	-	+	Substitution C/G	2	Protein M
28144	-	+	Substitution C/T	3	ORF8
28688	-	+	Substitution C/T	3	Protein N
28881	+	+	Substitution A/G	6 (O) / 2 (G)	Protein N
28882	+	+	Substitution A/G	5 (O) / 2 (G)	Protein N
28883	+	+	Substitution C/G	5 (O) / 2 (G)	Protein N

underscore the dynamic nature of SARS-CoV-2 evolution and its implications for public health. Phylogenetic analysis revealed significant genetic diversity among viral strains, with distinct subclass groupings observed within each country, our analysis led us to conclude that globally, the virus's distribution does not correlate with regions, lineages, or clades at the subclass level, underscoring the virus's immense diversity and mutability. Moreover, the high mutation rate and mutations present in clade O suggest a potential cause for the virus's reduced virulence.

By elucidating the genetic determinants of viral virulence, this study provides crucial insights that can inform the development of targeted therapeutic interventions and vaccine strategies aimed at combating COVID-19. Furthermore, the identification of specific mutations offers promising avenues for further investigation through molecular docking studies, which may unveil potential therapeutic targets for drug development. Moving forward, sustained surveillance efforts are imperative to monitor the ongoing evolution and transmission dynamics of SARS-CoV-2, facilitating timely interventions and control measures to curb the spread of the pandemic and minimize its impact on global health.

References

[1] Atzrodt CL, Maknojia I, McCarthy RD, Oldfield TM, Po J, Ta KT, et al. A Guide to COVID-19: A global pandemic caused by the novel coronavirus SARS-CoV-2. FEBS J. 2020;287(17):3633 -50. 10.1111/febs.15375

[2] Zatla I, Boublenza L, Hassaine H. SARS-CoV-2 morphology, genome, life cycle and our bodies' immune response: a review. Curr Top Virol. 2021. 18: 15-24 10.31300/CTVR.18.2021.15-24

[3] Ali SA, Baloch M, Ahmed N, Ali AA, Iqbal A. The outbreak of coronavirus disease 2019 (COVID-19)-An emerging global health threat. J Infect Public Health. 2020;13(4):644-6. 10.1016/j.jiph.2020.02.033

[4] Chatterjee, P., Nagi, N., Agarwal, A., Das, B., Banerjee, S., Sarkar, S., ... & Gangakhedkar, R. R. The 2019 novel coronavirus disease (COVID-19) pandemic: A review of the current evidence. The Indian journal of medical research, 2020. 151(2-3), 147.

[5] Valencia DN. Brief review on COVID-19: The 2020 pandemic caused by SARS-CoV-2. Cureus. 2020;12(3):e7386. 10.7759/cureus.7386

[6] El Zowalaty, M.E. and Järhult, J.D.. From SARS to COVID-19: A previously unknown SARS-related coronavirus (SARS-CoV-2) of pandemic potential infecting humans–Call for a One Health approach. One health, 2020. 9: 100124.

[7] Zatla I, Boublenza L, Hassaine H. SARS-CoV-2 origin, classification and transmission: a mini-review. Curr Top Virol. 2021. 18: 31-38. 10.31300/CTVR.18.2021.31-38

[8] Alanagreh L.A., Alzoughool F. and Atoum M.. The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms. Pathogens, 2020; 9(5), 331.

[9] Bonny, V., Maillard, A., Mousseaux, C., Plaçais, L., & Richier, Q. COVID-19: Pathophysiology of a Disease with Several Faces. The Internal Medicine Review. 2020.

[10] Saglietto, A., D'Ascenzo, F., Zoccai, G. B., & De Ferrari, G. M. COVID-19 in Europe: the Italian lesson. Lancet, 2020; 395 (10230): 1110-1111.

[11] Zatla I, Boublenza L, Boublenza A. Tracing the Origin and Early Progression of COVID-19 in Europe : An Epidemiological Descriptive Study. IEM 2023; 9 (3) :249-256. 10.61186/iem.9.3.249

[12] Zatla I., Boublenza L. The COVID-19 outbreak inception in Algeria: What Happened?. Algerian Journal of Health Sciences. 2023;18-15-24

[13] Martinez, I. L., Llinás, D. T., Romero, M. B., & Salazar, L. M. High Mutation Rate in SARS-CoV-2: Will It Hit Us the Same Way Forever. J Infect Dis Epidemiol. 2020; 6:176.

[14] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016; 33(7):1870-4. 10.1093/molbev/msw054.

[15] Talavera, G., & Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology. 2007.

[16] Marzin-Janvier, P. Supporting Experimental Design with Computer Environments: Studies in Genetics. RDST. Research in Didactics of Sciences and Technologies, 2015; 12: 87-112.

[17] Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., & Whittaker, G. R. Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically-sensitive activation loop. Journal of molecular biology. 2020.

[18] Suzuki, Y. Methods for Making Multiple Alignments of Genomic Sequences for Severe Acute Respiratory Syndrome Coronavirus 2. Meta Gene. 2020.

[19] Thieffry, D., & De Jong, H. Modeling, Analysis, and Simulation of Genetic Networks - Bioinformatics. Medicine/Sciences. 2002.

[20] Elayadeth-Meethal, M., Ollakkott S., Girish Varma G. COVID-19 and SARS-CoV-2: Molecular Genetics Perspectives. 2020.

[21] GISAID. Genomic epidemiology of novel coronavirus. 2020. Available at https://nextstrain.org/ncov

[22] Dao, T. L., Hoang, V. T., Colson, P., Lagier, J. C., Million, M., Raoult, D., Levasseur, A., & Gautret, P. SARS-CoV-2 Infectivity and Severity of COVID-19 According to SARS-CoV-2 Variants: Current Evidence. Journal of Clinical Medicine, 2021 10(12), 2635. 10.3390/jcm10122635

[23] Hamed, S. M., Elkhatib, W. F., Khairalla, A. S., & Noreddin, A. M. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. Scientific Reports. 2021;11(1), 1-8. 10.1038/s41598-021-87713-x