Iraqi Journal of Biotechnology, 2026, Vol. 25, No. 1, 187-194.

Co-existence and Dominancy of Mutational Cluster in the Genomes of SARS-CoV-2 In Iraq: Insights from Whole Genome Sequencing

[1] **Noor Saber,** [2] **Nuha J Kandala**

[1] National Center of Hematology, Mustansiriyah University, Iraq
[2] Department of Biotechnology, College of Science, University of Baghdad, Baghdad, Iraq

Received: October 9, 2024 Accepted: October 23, 2024 Published: March 30, 2026

Abstract

Background. The COVID-19 pandemic, caused by the SARS-CoV-2 virus, rapidly spread across the globe following its emergence in December 2019. SARS-CoV-2 exhibits significant genetic diversity, which has contributed to the evolution of various variants, including the Omicron sublineages. **Aim.** This study aims to investigate the genetic landscape of SARS-CoV-2 strains isolated from different governorates in Iraq during the Omicron wave, utilizing next-generation sequencing (NGS). **Methods.** A total number of 12 isolates collected and subjected to sequencing and bioinformatical analysis. **Results.** The results revealed the dominance of Omicron sublineages BA.1, BA.2, and their respective variants, along with 19 dominant genetic variations, including SNPs, deletions, and silent mutations. The current analysis highlights the persistence of different mutations in the RNA-dependent RNA polymerase, Nucleocapsid, and Membrane proteins. **Conclusion.** These findings underline the genetic variability and evolutionary adaptations of SARS-CoV-2 in Iraq, emphasizing the need for ongoing investigation and larger-scale studies to further understand the impact of these mutations on viral pathogenicity and immune response.

Keywords: Genomic Epidemiology, SARS-CoV-2, Next-Generation Sequencing, Iraq, Omicron

Corresponding author: (Email: noor.saber@uomustansiriyah.edu.iq)

Introduction

In December 2019, a group of newly emerged pneumonic infections reported from China. Later, on January 2020, Chinese authorities published the first sequence the causative agent which named SARS-CoV-2, within the one month, the virus spread over 28 countries, encompassing territories of WHO which ended with global declaration of COVID-19 pandemic (1). Despite the argument if the virus is airborne or aerosolized, the quarantine of COVID-19 patients a challenge to public health mitigations strategies as most of the medical approaches include aerosol-generating equipment including the open suctioning, nebulizers, and endotracheal intubation (1,2).

Coronaviruses are an enveloped, positively sense RNA viruses that target diverse animals hosts. They belong to the subfamily Orthocoronavirinae, Corona-viridae family. Coronaviruses fall within four genera: Alpha, Beta, Gamma, and Delta. Alpha and Beta coronaviruses target mammalian hosts, while Gamma and Delta coronaviruses mostly infect birds. These viruses derive their envelope from the host cell membrane, which incorporates viral surface proteins. The spike protein, which possess the binding area of targeted host

receptors, are protruding from viral membrane give them the appearance of crown under microscopes (1,2).

The positive viral characteristic allows the virus to act directly as mRNA and translated into proteins via host protein synthesis machinery. SARS-CoV-2 share 80% genomic identity with SARS-CoV-1, ranging in size ranging from 60 to 140 nm. The genomes include approximately 29,891 nucleotides coding for 9860 amino acids resembled in four structural proteins: Spike (S), Envelope (E), Nucleocapsid (N), and Membrane (M), in addition to 16 non-structural proteins. These are coded by various open reading frames, primarily the orf1ab (1,2). In 2009, the PREDICT program, a part of USAID's Emerging Pandemic Threats initiative, implement genomic epidemiology strategy as part of advanced surveillance plan that led to the identification of over 931 novel viruses, such efforts highlighted the effectiveness of molecular techniques such as sequencing in epidemiological studies (3,4). These facts highlight the need for implementation of genomic epidemiological techniques in biological research.

Globally, NGS is employed as a part of public health diagnostic systems that aid in the monitoring, tracking, and surveillance of circulating or emerging pathogens (5,6). The current study conducts a whole genome paired end (forward and reverse) next generation sequencing (NGS) using Illumina MiSeq sequencer to investigate the whole genome of SARS-CoV-2 virus isolated from different Iraqi governorates.

Materials and methods

A total number of twelve nasopharyngeal specimens collected from different governorates during Omicron 4th infectious wave, these specimens distributed as following, six from Baghdad, three from Missan, One from Karbala, one from Diyala and one from ThiQar.

RNA extraction and RT-PCR Assay

RNA extraction procedure considered using automated Bioneer ExiPrep™ 96 Lite extraction machines (A-5250, BIONEER), the extraction kit that used for the study is the Exiprep 96 Viral DNA/RNA kit (K-4614, BIONEER). The extracted RNA was used as a template for RT-PCR assay to detect the viral load and confirm the positivity for SARS-CoV-2 infection. The RT-PCR assay that is used is the TaqPath assay. This assay is a commercial kit that includes all the content required to perform the PCR reaction which is supplied by Thermo Fisher. The used RT-PCR machine for the assay is the Applied Biosystem 7500 fast system.

Sequencing Workflow

To ensure a robust analysis of the genetic diversity and to achieve a deeper comprehensive insight into the dynamic evolutionary nature of SARS-CoV-2 during the studied timeframe (7,8). The libraries were prepared using the Illumina AmpliSeq protocol (Illumina, USA), which involved four key steps: cDNA preparation, target amplification (SimpliAmp Thermal Cycler, Applied Biosystem), library preparation, and sequencing (9). All kits that are used to perform the sequencing by synthesis (SBS) reaction starting from the first step till ending with library elution through the universal AmpliSeq workflow are listed in Table 1. The final step is sequencing using MiSeq Sequencer (Illumina, USA). The entire procedure is followed as recommended by the manufacturer instruction without modifications.

Table 1: List of contents of AmpliSeq Illumina kits compatible with MiSeq instrument.

Kit Description	Content	Reference Number	Brand
AmpliSeq cDNA Synthesis for Illumina	5X AmpliSeq cDNA Reaction Mix	20022654	Illumina
	10X AmpliSeq RT Enzyme Mix		
AmpliSeq Library PLUS for Illumina	1X Lib Amp Mix	20019102	Illumina
	10X Library Amp Primers		
	DNA Ligase		
	5X AmpliSeq HiFi Mix		
	FuPa Reagent		
	Low TE*		
	Switch Solution		
AmpliSeq Library Equalizer for Illumina	Equalizer Beads	20019171	Illumina
	Equalizer Capture		
	Equalizer Elution Buffer		
	Equalizer Wash Buffer		
	Equalizer Primer		
MiSeq V2 Reagent Kit	Reagent Cartridge	15033626	Illumina
	H1T1 Hybridization Buffer		
	PR2 Incorporation Buffer	15033624	Illumina
	Flow Cell		
AmpliSeq™ for Illumina® Custom RNA Panel	AmpliSeq custom RNA 2&1 primers set panel	20020496	Illumina
AmpliSeq™ CD Indexes Set A for Illumina	(96 Indexes, 96 Samples)	20019105	Illumina

Analysis Pipeline

The resulted FASTQ files subjected to quality checking using FASTQC software which is a standard and powerful tool for quality control procedures of NGS data (10) . All generated FASTQ files trimmed to remove sequencing artifacts resembled by adapters and low-quality reads from FASTQ files using reads using BBDUK tool (11). Resulted trimmed paired end sequencing files merged and assembled to a reference using map to reference assembly powered by Geneious Prime (12), the selected retrieved from NCBI RefSeq database along with the annotation, the reference designated with the accession number NC_045512.2. Single nucleotide polymorphisms (SNPs), substitutions that include more than one mutation in adjacent cites, insertions and indels identified using Variations/SNPs finder tool powered by Geneious prime. Conversion of FASTQ to FASTA formats considered using Geneious Prime. Variant identification was conducted using Nextclade platform (13). Phylogenetic tree construction considered using MegaX (14), and Auspice platform (15). Alignment of generated FASTA file considered using BLASTn/NCBI and Audacity Instant powered by Global Initiative on Sharing All Influenza Data, GISAID (16–18). All raw FASTQ data uploaded to short read archive (SRA) on NCBI.

Data Availability

All sequencing data available in row format on short read archive (SRA) on NCBI, the accession numbers listed in table 2.

Table 2: The sequencing data available in row format on short read archive (SRA) on ncbi

Samples ID	Bio project Accession ID	Bio sample accession	Accession ID	Province	Collection Date
1	PRJNA913479	SAMN32299962	SRR22801243	Missan	3/23/2022
2	PRJNA913479	SAMN32299963	SRR22801242	Missan	3/23/2022
3	PRJNA913479	SAMN32299964	SRR22801239	Baghdad	3/23/2022
4	PRJNA913479	SAMN32299965	SRR22801238	Baghdad	3/23/2022
5	PRJNA913479	SAMN32299966	SRR22801237	Baghdad	3/23/2022
6	PRJNA913479	SAMN32299967	SRR22801236	Karbala	3/28/2022
7	PRJNA913479	SAMN32299968	SRR22801235	ThiQar	3/29/2022
8	PRJNA913479	SAMN32299969	SRR22801234	Baghdad	3/29/2022
9	PRJNA913479	SAMN32299970	SRR22801233	Baghdad	3/29/2022
11	PRJNA913479	SAMN32299971	SRR22801232	Missan	7/4/2022
11	PRJNA913479	SAMN32299972	SRR22801241	Diyala	7/4/2022
12	PRJNA913479	SAMN32299973	SRR22801240	Baghdad	7/4/2022

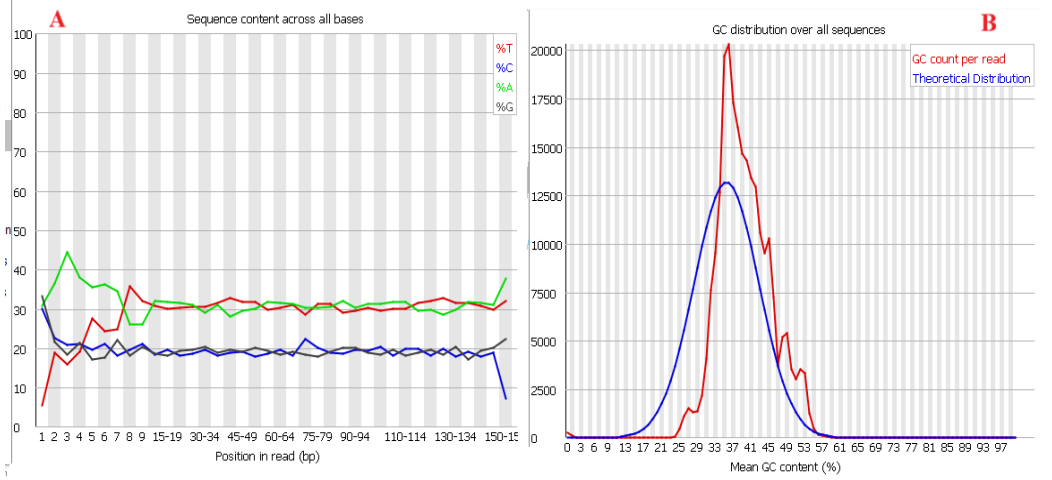

Figure 1: GC% content of analyzed FASTQ files, A: Sequence content across all bases, B: GC distribution over all sequences, blue curve represents the normal theoretical distribution of bases, red curve represents the GC count

The mean assembled genome length was 29,740 bp, spanning nucleotide positions 31 to 29,858 when mapped to the Wuhan reference genome (NC_045512.2). Gaps were primarily localized within the 5′ and 3′ untranslated regions (UTRs). A total of 25,654 sites (86.0%) were identical across the analyzed sequences (21). Alignment analysis revealed mismatches relative to the reference genome, corresponding to single nucleotide polymorphisms (SNPs), substitutions, and deletions. Variant assignment using Nextclade identified all isolates as belonging to the Omicron lineage. Sublineages included BA.1, BA.1.1, BA.1.17.2, BA.2, and BA.2.12, within clades 21K and 21L (Figure 2).

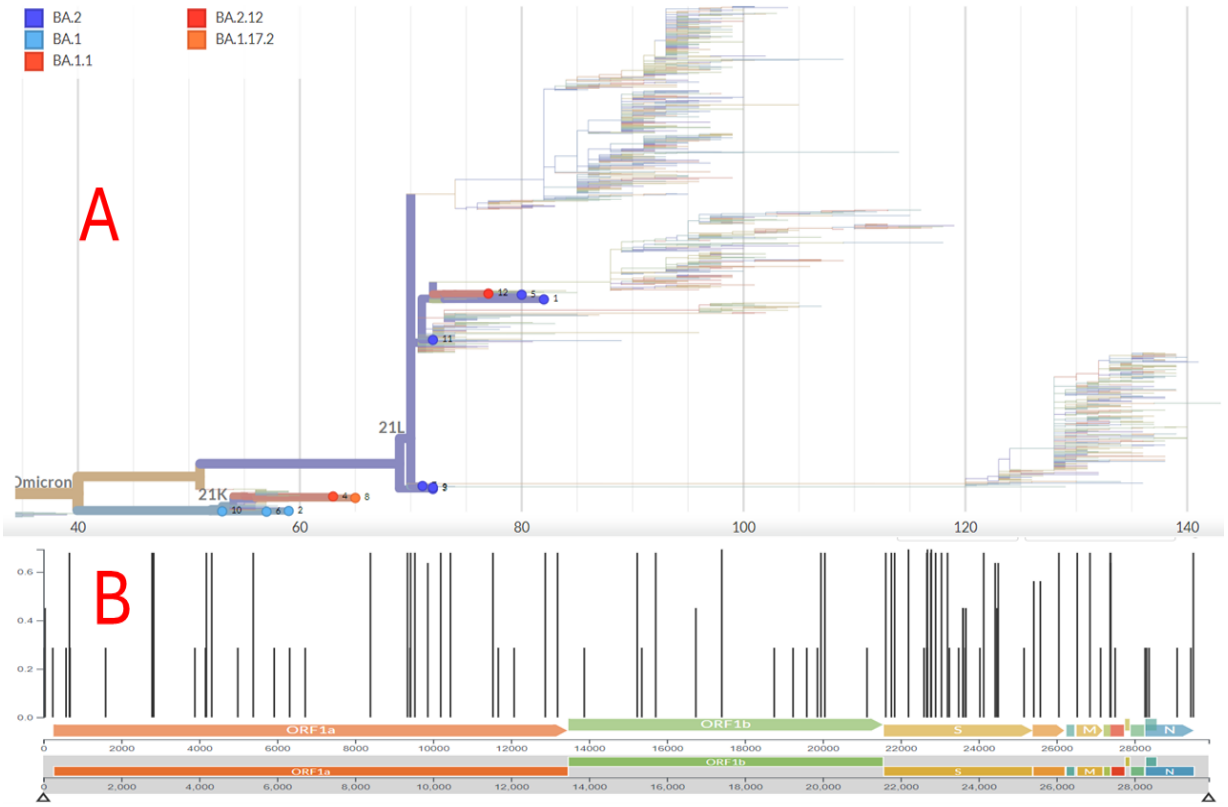

Figure 2: SARS-CoV-2 variant prediction, A: dominancy of BA.2 sublinages. Furthermore, the results indicate the presence of another variant like BA.1, BA.1.1, BA.2.12 and BA.1.17.2. These sublineages fall within clades 21K and 21L of the Omicron evolutionary tree. B: Show the locations of harbored variations across ORF1a, ORF1b, Spike, Membrane and Nucleocapsid genes.

Comprehensive mutation analysis identified 19 dominant co-existing variations across all isolates. These include 13 SNPs, 2 deletions, 1 substitution resulting from two adjacent SNPs in the same codon, and 3 silent mutations. The detailed list of mutations is presented in Table 3.

Table 3: Dominant and co-existed genetic variations that were detected across all analyzed isolates.

Variation ID	Changed Codon	Location	Variation Type
T3225I	ACC→ATC	ORF1ab	SNP
P3395H	CCC→CAC	ORF1ab	SNP
3675-3677 SGF deletion	TCT,GGT,TTT deletion	ORF1ab	Deletion of three codons
P4715L	CCT→CTT	ORF1ab (RdRp)	SNP
I5964V	GTA→ATA	ORF1ab	SNP
G339D	GGT→GAT	Spike	SNP
K417N	AAG→AAT	Spike	SNP
D614G	GAT→GGT	Spike	SNP
H655Y	CAT→TAT	Spike	SNP
D796Y	GAT→TAT	Spike	SNP

D1146	**GAC→GAT**	**Spike**	**Silent Mutation**
T9I	**ACA→ATA**	**Envelope**	**SNP**
Q19E	**CAA→GAA**	**Envelope**	**SNP**
A63T	**GCT→ACT**	**Membrane**	**SNP**
R20	**AGG→CGG**	**ORF6**	**Silent Mutation**
L18	**CTA→TTA**	**ORF7b**	**Silent Mutation**
31-33 ERS deletion	**A,GAA,CGC,AG, deletion**	**Nucleocapsid**	**Deletion of three codons**
R203K	**AGG→AAA**	**Nucleocapsid**	**Substitution, two adjacent SNPs in the same codon**
G204R	**GGA→CGA**	**Nucleocapsid**	**SNP**

Among ORF1ab mutations, P4715L (RdRp region) was consistently detected. Additional mutations were identified across the Spike, Envelope, Membrane, ORF6, ORF7b, and Nucleocapsid genes. The Nucleocapsid gene harbored the double mutation R203K and G204R, as well as a three-codon deletion (31–33 ERS deletion).

Discussion

The observed GC distribution aligns with the intrinsic genomic composition of SARS-CoV-2 (~38% GC), which explains the deviation from standard theoretical GC curves (20). The 86% sequence identity across isolates reflects high genomic conservation despite the presence of mutations. Such variability is a hallmark of RNA viruses and contributes to the emergence of new variants (22,23). The elevated mutation burden observed in the Spike gene supports previous findings indicating accelerated evolution of this region compared to the rest of the viral genome (24–27). The predominance of Omicron sublineages BA.1 and BA.2 is consistent with global epidemiological trends during the study period (28–30). Omicron's rapid expansion has been attributed to enhanced immune evasion and infectivity driven by accumulated mutations (28–30). The consistent co-existence of 19 dominant mutations suggests the presence of a stable mutational signature within circulating Iraqi isolates. Recurrent mutation patterns are recognized drivers of viral adaptation across geographic regions (31). The persistence of P4715L in the RdRp region is notable due to its location within the viral replication machinery. This mutation has previously been associated with epitope loss and potential immune evasion, although it is considered functionally neutral (32). P4715L was first reported in Italy and subsequently observed at high frequency across multiple continents (33), indicating global dissemination.

The Envelope T9I mutation has recently been associated with resistance to autophagy (34), which may influence viral-host interactions. Mutations within the N and M proteins are also biologically relevant, given their involvement in immune modulation and viral assembly (35). The R203K/G204R double mutation in the Nucleocapsid protein has been shown to enhance viral replication, increase RNA levels, and promote phosphorylation-dependent mechanisms that augment viral fitness (36). Similar mutations in N and M genes have been previously reported in Iraqi isolates (37). Although the current dataset provides insight into circulating mutational patterns, larger-scale studies are necessary to validate these

observations and assess their phenotypic impact.

Conclusion

The persistence and spread of mutations like P4715L in the RdRp gene and T9I in the Envelope protein highlight the evolutionary adaptations of the virus, potentially contributing to its enhanced infectivity and immune evasion capabilities. These findings underscore the importance of continuous genomic surveillance to track viral evolution and to better understand the implications of these genetic changes on public health. Future studies with larger sample sizes and further delve into the mutational effect are essential to confirm these findings and to provide further insights into the mutational dynamics of SARS-CoV-2 in Iraq. Such efforts are crucial for informing public health strategies and for developing targeted interventions to manage and mitigate the impact of emerging variants.

Acknowledgment

The authors thank Central Public Health Laboratory for their support.

Ethical Approval

The research topic approved by the ethics committee at College of Science, University of Baghdad (CSEC/0222/0155) on January 28th, 2022.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Funding

The current work didn't receive funds or grants from public or private agencies.

References

1. Ludwig, S. and Zarbock, A. (2020). Coronaviruses and SARS-CoV-2: A Brief Overview. Anesthesia and Analgesia, 131(1), 93–96.
2. Shi, Y.; Wang, G.; Cai, X.; Deng, J.; Zheng, L.; Zhu, H. et al. (2020). An overview of COVID-19. Journal of Zhejiang University: Science B, 21(4), 343–360.
3. Kelly, T. R.; Machalaba, C.; Karesh, W. B.; Crook, P. Z.; Gilardi, K.; Nziza, J. et al. (2020). Implementing One Health approaches to confront emerging and re-emerging zoonotic disease threats: lessons from PREDICT. One Health Outlook, 2(1), 1.
4. Carlson, C. J. (2020). From PREDICT to prevention, one pandemic later. Lancet Microbe, 1(1), e6–e7.
5. Gwinn, M.; Maccannell, D. and Armstrong, G. L. (2019). Next-Generation Sequencing of Infectious Pathogens. JAMA - Journal of the American Medical Association, 321(9), 893–894.
6. Papa Mze, N.; Kacel, I.; Beye, M.; Tola, R.; Sarr, M.; Basco, L. et al. (2023). High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. Genes (Basel), 14(3).
7. Satam, H.; Joshi, K.; Mangrolia, U.; Waghoo, S.; Zaidi, G.; Rawool, S. et al. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. Biology, 12(3).
8. Grubaugh, N. D.; Ladner, J. T.; Lemey, P.; Pybus, O. G.; Rambaut, A.; Holmes, E. C. et al. (2019). Tracking virus outbreaks in the twenty-first century. Nature Microbiology, 4(1), 10–19.
9. Papa Mze, N.; Kacel, I.; Beye, M.; Tola, R.; Sarr, M.; Basco, L. et al. (2023). High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. Genes (Basel), 14(3).
10. Trivedi, U. H.; Cézard, T.; Bridgett, S.; Montazam, A.; Nichols, J.; Blaxter, M. et al. (2014). Quality control of next-generation sequencing data without a reference. Frontiers in Genetics, 5(MAY).
11. Singer, E.; Andreopoulos, B.; Bowers, R. M.; Lee, J.; Deshpande, S.; Chiniquy, J. et al. (2016). Next generation sequencing data of a defined microbial mock community. Scientific Data, 3.
12. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S. et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647–1649.
13. Hadfield, J.; Megill, C.; Bell, S. M.; Huddleston, J.; Potter, B.; Callender, C. et al. (2018).

NextStrain: Real-time tracking of pathogen evolution. Bioinformatics, 34(23), 4121–4123.
14. Tamura, K.; Stecher, G. and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution, 38(7), 3022–3027.
15. Aksamentov, I.; Roemer, C.; Hodcroft, E. and Neher, R. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software, 6(67), 3773.
16. Elbe, S. and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges, 1(1), 33–46.
17. Khare, S.; Gurry, C.; Freitas, L.; Schultz, M. B.; Bach, G.; Diallo, A. et al. (2021). GISAID's Role in Pandemic Response. China CDC Weekly, 3(49), 1049–1051.
18. Shu, Y. and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance, 22(13).
19. Cock, P. J. A.; Fields, C. J.; Goto, N.; Heuer, M. L. and Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research, 38(6), 1767–1771.
20. Fumagalli, S. E.; Padhiar, N. H.; Meyer, D.; Katneni, U.; Bar, H.; DiCuccio, M. et al. (2023). Analysis of 3.5 million SARS-CoV-2 sequences reveals unique mutational trends with consistent nucleotide and codon frequencies. Virology Journal, 20(1).
21. Cao, C.; Cai, Z.; Xiao, X.; Rao, J.; Chen, J.; Hu, N. et al. (2021). The architecture of the SARS-CoV-2 RNA genome inside virion. Nature Communications, 12(1).
22. Carabelli, A. M.; Peacock, T. P.; Thorne, L. G.; Harvey, W. T.; Hughes, J.; de Silva, T. I. et al. (2023). SARS-CoV-2 variant biology: immune escape, transmission and fitness. Nature Reviews Microbiology, 21(3), 162–177.
23. Markov, P. V.; Ghafari, M.; Beer, M.; Lythgoe, K.; Simmonds, P.; Stilianakis, N. I. et al. (2023). The evolution of SARS-CoV-2. Nature Reviews Microbiology, 21(4), 361–379.
24. Greaney, A. J.; Loes, A. N.; Crawford, K. H. D.; Starr, T. N.; Malone, K. D.; Chu, H. Y. et al. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe, 29(3), 463–476.
25. Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J. C. C. et al. (2020). Escape from neutralizing antibodies 1 by SARS-CoV-2 spike protein variants. Elife, 9. 26. Perez-Gomez, R. (2021). The Development of SARS-CoV-2 Variants: The Gene Makes the Disease. Journal of Developmental Biology, 9(1).
27. Bar-On, Y. M.; Flamholz, A.; Phillips, R. and Milo, R. (2020). Sars-cov-2 (Covid-19) by the numbers. Elife, 9.
28. Willett, B. J.; Grove, J.; MacLean, O. A.; Wilkie, C.; De Lorenzo, G.; Furnon, W. et al. (2022). SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. Nature Microbiology, 7(8), 1161–1179.
29. Chen, L.; He, Y.; Liu, H.; Shang, Y. and Guo, G. (2024). Potential immune evasion of the severe acute respiratory syndrome coronavirus 2 Omicron variants. Frontiers in Immunology, 15.
30. Ao, D.; He, X.; Hong, W. and Wei, X. (2023). The rapid rise of SARS-CoV-2 Omicron subvariants with immune evasion properties: XBB.1.5 and BQ.1.1 subvariants. MedComm, 4(1).
31. van Dorp, L.; Acman, M.; Richard, D.; Shaw, L. P.; Ford, C. E.; Ormond, L. et al. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infection, Genetics and Evolution, 83.
32. Gupta, A. M.; Chakrabarti, J. and Mandal, S. (2020). Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants. Microbes and Infection, 22(10), 598–607.
33. Omotoso, O. E.; Olugbami, J. O. and Gbadegesin, M. A. (2021). Assessment of intercontinents mutation hotspots and conserved domains within SARS-CoV-2 genome. Infection, Genetics and Evolution, 96.
34. Klute, S.; Nchioua, R.; Cordsmeier, A.; Vishwakarma, J.; Koepke, L.; Alshammary, H. et al. (2024). SARS-CoV-2 Omicron Envelope T9I adaptation confers resistance to autophagy. bioRxiv.
35. Hossain, A.; Akter, S.; Rashid, A. A.; Khair, S. and Alam, A. S. M. R. U. (2022). Unique mutations in SARS-CoV-2 Omicron subvariants' non-spike proteins: Potential impacts on viral pathogenesis and host immune evasion. Microbial Pathogenesis, 170.
36. Johnson, B. A.; Zhou, Y.; Lokugamage, K. G.; Vu, M. N.; Bopp, N.; Crocquet-Valdes, P. A. et al. (2022). Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. PLoS Pathogens, 18(6).
37. Kradi, A. M.; Hussein, A.; Al-Faisal, M. and Turki, A. M. (2022). Alteration of M and N Genes of Corona Virus 2 (SARS-CoV-2). Iraqi Journal of Biotechnology, 21