Virus Research 345 (2024) 199392

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

A systematic mutation analysis of 13 major SARS-CoV-2 variants

Han Bai [a,1], Xuan Zhang [b,c,d,e,1], Tian Gong [b,c,d,e,1], Junpeng Ma [b,c,d,e], Peng Zhang [b,c,d,e], Zeqiong Cai [a], Doudou Ren [a], Chengsheng Zhang [a,b,c,d,e,*]

[a] The MED-X Institute, The First Affiliated Hospital of Xi'an Jiaotong University, Building 21, Western China Science and Technology Innovation Harbor, Xi'an 710000, China
[b] Center for Molecular Diagnosis and Precision Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 1519 Dongyue Dadao, Nanchang 330209, China
[c] Department of Clinical Laboratory, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 17 Yongwai Zhengjie, Nanchang 330006, China
[d] Jiangxi Provincial Center for Advanced Diagnostic Technology and Precision Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 1519 Dongyue Dadao, Nanchang 330209, China
[e] Department of Medical Genetics, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 1519 DongYue Dadao, Nanchang 330209, China

ARTICLE INFO

Keywords:
SARS-CoV-2
Variants
Mutation
Spike protein
Sequence analysis
Molecular dynamics analysis

ABSTRACT

SARS-CoV-2 evolves constantly with various novel mutations. Due to their enhanced infectivity, transmissibility and immune evasion, a comprehensive understanding of the association between these mutations and the respective functional changes is crucial. However, previous mutation studies of major SARS-CoV-2 variants remain limited. Here, we performed systematic analyses of full-length amino acids mutation, phylogenetic features, protein physicochemical properties, molecular dynamics and immune escape as well as pseudotype virus infection assays among thirteen major SARS-CoV-2 variants. We found that Omicron exhibited the most abundant and complex mutation sites, higher indices of hydrophobicity and flexibility than other variants. The results of molecular dynamics simulation suggest that Omicron has the highest number of hydrogen bonds and strongest binding free energy between the S protein and ACE2 receptor. Furthermore, we revealed 10 immune escape sites in 13 major variants, some of them were reported previously, but four of which (i.e. 339/373/477/496) are first reported to be specific to Omicron, whereas 462 is specific to Epslion. The infectivity of these variants was confirmed by the pseudotype virus infection assays. Our findings may help us understand the functional consequences of the mutations within various variants and the underlying mechanisms of the immune escapes conferred by the S proteins.

1. Introduction

The outbreak of severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) caused a global pandemic in 2020 and led to millions of deaths across the world (WHO Coronavirus (COVID-19) Dashboard 2020; World Health Organization, 2021). Moreover, the emergence of diverse SARS-CoV-2 variants with enhanced transmissibility, pathogenicity, and decreased efficacy of medical countermeasures has exacerbated global concerns. Based on the transmissibility, disease severity, risk of reinfection, and impacts on diagnostics and vaccine performance, Alpha, Beta, Gamma, Delta, and Omicron have been listed as variants of concern (VOC) whereas Lambda and Mu have been listed as variants of interest (VOI) by World Health Organization (WHO) (Li et al., 2020; Arunachalam et al., 2021; Campbell et al., 2021). Statistical analysis reveals that the Delta variant has 15 mutations in the S protein compared to the original strain, while the Omicron variant has accumulated 37 mutations in the S protein (Telenti et al., 2022). Currently, Omicron has become the dominant variant globally, accounting for 98 % of infections. Its latest sub-lineages, including BA.3, BA.4, BA.5, BA.2.12.1, BF.7 and XBB, have demonstrated higher transmissibility than BA.2 and are under continuous surveillance by the WHO (Xia et al., 2022; Sarkar et al., 2023).

SARS-CoV-2 is an enveloped, plus-stranded RNA virus (Wu et al., 2020). Due to the high likelihood of mutations during the replication

* Corresponding author at: Center for Molecular Diagnosis and Precision Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 1519 Dongyue Dadao, Nanchang 330209, China.
E-mail address: ndyfy09564@ncu.edu.cn (C. Zhang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.virusres.2024.199392
Received 24 December 2023; Received in revised form 22 April 2024; Accepted 7 May 2024
Available online 15 May 2024

process of single-stranded RNA, the mutations generated during the viral replication cannot be promptly corrected, subsequently leading to viral mutations. Of note, coronaviruses encode an exonuclease proofreading-repair enzyme, but it is unknown to what extent the exonuclease is working for SARS-CoV-2. While the mutation tracking studies have shown that SARS-CoV-2 mutations are evenly distributed throughout the entire viral genome of approximately 29 kb size, the preservation of these mutations is not random but driven by a multitude of competitive processes, such as host editing at the organismal level induced by immune responses and host-virus interactions, as well as natural selection response to environmental changes or adaptations to new conditions at the population level (Foster, 2000; Metzgar and Wills, 2000). Studies on long-term COVID-19 cases have found that compromised immune function is a key factor in persistent infection, and long-term infected individuals may also serve as a source for new variants (Goyal et al., 2021).

SARS-CoV-2 genome is composed of a total of 11 genes with 11 open reading frames (ORFs). Approximately two-thirds of the genome is at the 5′ terminal encoding the nonstructural protein ORF1ab, which functions as an mRNA to translate two polyproteins (ORF1a and ORF1b) cleaved by proteases into sixteen nonstructural proteins (nsp1–16) to initiate viral genome replication and transcription (Yadav et al., 2021). The remaining one-third viral genome is at the 3′ terminal encoding structural proteins including, spike glycoprotein (S), membrane (M), envelope protein (E), nucleocapsid (N), and accessory proteins of ORF3a, ORF6, ORF7, ORF8 and ORF10 (Yoshimoto, 2020). The S consists of S1 and S2 subunits. The S1 subunit includes the N-terminal domain (NTD) and receptor binding domain (RBD) that contains the receptor binding motif (RBM), and furin cleavage site (cleave S1 and S2). The S2 subunit includes the fusion peptide (FP), connecting domain (CD), heptad repeat domain (HRD), transmembrane domain (TM), and cytoplasmic tail (CT) (Kirchdoerfer et al., 2018; Cai et al., 2020). Because of the critical roles of S protein in mediating receptor recognition, cellular entry, viral infection and immune escape, the mutations in S bear unique signatures of both purifying selection and the adaptation of SARS-CoV-2 to humans (Thorne et al., 2022).

A study quantifying adaptability and mutation accumulation based on early SARS-CoV-2 genomic data found a positive correlation between the number of non-synonymous mutations in the S1 subunit and the increased growth rate of the viral strain (Kistler et al., 2022). More specifically, the mutations at the ACE2-RBD interface influence the binding affinity of SARS-CoV-2 with the ACE2 receptor by changing salt bridges, hydrogen bonds, van der Waals, ionic interactions, and solvent-accessible surface area (Lan et al., 2020). For instance, new hydrogen bonds formed by mutated residues Q493R, G496S, Q498R, and N501Y in the RBD with ACE2 enhance ACE2 binding affinity (Mannar et al., 2022). K417N mutation abolishing multiple hydrogen bonds and salt-bridge hydrogen bonding form pairs reduce ACE2 binding affinity (Qu et al., 2023). In addition, the mutations in S cause physicochemical alterations in the SARS-CoV-2 spike protein attributing additional advantages to the virus. For instance, there is an increase in the positive electrostatic potential on the RBD surface of the variants with N440K mutation, due to the positive charge of the lysine residue, which leads to a stronger binding of the RBD-hACE2 complex than the wild type SARS-CoV-2 (Rani et al., 2021).

The S477N substitution located at the protein flexible region border has been suggested to increase the local conformation stability and confer a greater rigidity to the protein base (Izumi et al., 2020). It has been recognized that the greater the rigidity caused by the mutation, the more stable the local conformation of the protein becomes (Ponde, 2023). Therefore, a mutation that promotes greater conformational rigidity favors protein expression and viral infectivity. Furthermore, when the physicochemical properties were changed by the mutations occurred in the antigenic sites, they can result in a decrease in the neutralizing capacity of previously generated antibodies from past infections or vaccination, rendering the viruses incompletely cleared and causing immune evasion. For instance, functional studies have shown that the S with E484K/Q/A mutation significantly reduced the neutralizing effect of serum antibodies from recovered COVID-19 patients (Cosar et al., 2022), and the S with L452R mutation led to significant immune evasion against RBD antibodies (Desingu and Nagarajan, 2022). The S with D614G mutation increased conformational flexibility and could contribute to decreased neutralizing antibody recognition and enhanced viral replication and transmission capabilities (Zhou et al., 2021; de Souza, 2022).

Moreover, based on quantifying mutations, the adaptive evolution in multiple regions of the coronavirus genome has been partially characterized (Kistler et al., 2022). Vipul et al. found Alpha (N501Y), Kappa (L452R, E484Q), and Delta (L45R, T478K) succeeded in enhancing binding hACE2 with the receptor-binding domain (RBD) when compared with the wild type (WT). In the Kappa and Delta variants, the L452R, T478K, and E484Q mutations increase the stability of the spike protein and intra-chain interactions, potentially weakening the interaction between neutralizing antibodies and the strains (Mlcochova et al., 2021). Additionally, Bloom and colleagues have scanned all possible combinations of amino acid mutations at the 201 residues of the receptor-binding domain (RBD) and examined their effects on the binding ability to hACE2 and monoclonal or serum antibodies using a yeast system, and suggested that Omicron may evolve mutations to evade the humoral immunity (Greaney et al., 2021; Muecksch et al., 2021).

Although previous studies indicated that these mutations may significantly change the infectivity and immunological characteristics of the variants, the relationship between physicochemical properties, biological consequences, evolution, and new mutations in existing variants has not been systematically and comprehensively investigated (Chakraborty et al., 2021). Most of the previous mutation studies focused on the spike protein, or Omicron and Delta, or variants in specific countries and regions, or only amino acid (aa) residue substitutions excluding aa deletions and insertions. Therefore, a more comprehensive and systematic mutation analysis is urgent needed. In this study, we combined the silicon analysis and experimental studies, and systematically performed the sequence and structure alignments, molecular dynamics analysis and immune escape analysis among thirteen major SARS-CoV-2 variants, with emphasis on Omicron and Delta. We hope that our study may provide novel insights into understanding the impact of various mutations on the pathogenicity of SARS-CoV-2 and developing effective strategies against the ongoing infections of SARS-CoV-2.

2. Materials and methods

2.1. Plasmids, cell lines and cell culture

Codon-optimized cDNA encoding S glycoprotein of SARS-CoV-2 Wuhan-1 (NC_045512), Kappa (MZ571142), Delta (MZ377115) and Omicron (OL672836), was synthesized and cloned into eukaryotic cell expression vector pcDNA3.1(+) or pCMV3 (TSINGKE, China). The lentiviral packaging plasmid psPAX2 was obtained from Huayueyang, China. The pLenti-GFP-luci lentiviral reporter plasmid that expresses GFP and luciferase was purchased from TSINGKE. The pLVX-puro-ACE2 plasmids were also synthesized directly by TSINGKE. All constructs in this study were confirmed by DNA sequencing.

The HEK293T (293T) cell line (human embryonic kidney cells) was maintained in Dulbecco's Modified Eagle's Medium (SIGMA, Germany), supplemented with 10 % fetal bovine serum (FBS, Biological Industries, America) and 1 % penicillin-streptomycin 100× Solution (PS, Hyclone, America). Cell cultures were maintained in a humidified incubator at 37 °C in 5 % CO_2 in the indicated media and passaged every 3–4 days. The 293T stable cell lines expressing human ACE2 (293T/hACE2) were established in our lab with the following method: Lentivirus packaging vectors psPAX2 and pLenti-GFP-luci were co-transfected with pLVX-puro or pLVX-puro-ACE2 plasmid using Lipo8000 (Beyotime, China)

in 293T cells. At 48 h post transfection, puromycin was added to a final concentration of 3 μg/mL. The cells were selected for 7 days to maintain the stable expression of the empty vector or ACE2, respectively.

2.2. Sequence alignment and phylogenetic analysis

A total of 130 complete coding sequences (CDS) of thirteen major SARS-CoV-2 variants associated with the COVID-19 pandemic were retrieved stochasticly from NCBI Virus (https://www.ncbi.nlm.nih.gov/labs/virus/) for the sequence alignment analysis with the reference sequence Wuhan-Hu-1 (NCBI: NC_045512) (Table 1) (Hatcher et al., 2017). In addition, 3216 sequences from the GISAID database (https://www.gisaid.org) were processed for phylogenetic analysis (Shu and McCauley, 2017; Hadfield et al., 2018). For sequence alignment, we employed the multiple sequence alignment and phylogenetic analysis tool Clustalw (https://www.genome.jp/tools-bin/clustalw) and COBALT (https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?LINK_LOC=BlastHomeLink) (Hung and Weng, 2016). For the phylogenetic analysis, we utilized the Phylogeny tool (https://www.gisaid.org/phylodynamics/global/nextstrain/) (Wang et al., 2020).

2.3. Homology modelling, structure alignment and physicochemical property analysis

The SARS-CoV-2 S protein structure QHD43416 from I-TASSER (https://www.zhanggroup.org/COVID-19/) based on experimental structure (PDB ID: 6VYB, 6VXX, 6LXT) was used as template (Roy et al., 2010). Homology models of SARS-CoV-2 variants were built based on the website SwissModel (https://www.swissmodel) (Waterhouse et al., 2018). We uploaded the template protein and amino acid sequence of each variant with default parameters, respectively. SwissModel will calculate and generate a 3D protein model automatically. In detail, it first transfers conserved atom coordinates as defined by the target-template alignment. Then the residue coordinates corresponding to insertions/deletions in the alignment are generated by loop modelling and a full-atom protein model is obtained by constructing the non-conserved amino acid side chains. SwissModel relies on the OpenStructure computational structural biology framework and the ProMod3 modelling engine to perform this step [40]. The structure alignment has been performed and displayed by PyMOL (Lill and Danielson, 2011). The physicochemical property was analyzed via Expasy (https://www.expasy.org/). We uploaded the amino acid sequence of each variant spike protein and select the different modules in website, such as pI, instability index, aliphatic index, and grand average of hydropathy (GRAVY) of the proteins. Then with default parameters, Expasy ProtParam tool will compute and output corresponding result. The principal component analysis (PCA) was performed using omicstudio (https://www.omicstudio.cn/tool/144).

2.4. Molecular dynamics (MD) simulations

Pretreatment steps such as small molecule (Zn, CL) removal, hydrogenation, hydrogen bond optimization and restriction minimization were performed on the structures, respectively. Topologies of all systems were neutralized with Na+/Cl and solvated using the TIP3P water model. The systems were embedded in a rectangular water box with an edge distance of 10 Å. Molecular dynamics simulations were carried out using the GROMACS 2022 with the amber99sb-ildn force field (Van Der Spoel et al., 2005). The Brownian motion of the system was minimized by 100 ps, and then equilibration was performed with constant volume and temperature (NVT) and constant pressure and temperature (NPT). The unrestricted dynamic simulation was performed for 50 ns under the NPT of 300 K and 1atm. The trajectory analysis was performed using GROMACS tools, the parameters root mean square deviation (RMSD), root mean square fluctuation (RMSF), and the amount of H-bonds were all included, and finally trajectories were visualized by Xmgrace. Six

Table 1
Classification and first reported region of the wild type and thirteen variants of SARS-CoV-2.

WHO ID	Wild type and 13 variants of SASR-CoV-2														
	Wild type	Alpha α	Beta β	Gamma γ	Delta δ	Omicorn ο	Epsilon ε	Zeta ζ	Eta η	Theta θ	Iota ι	Kappa κ	Lambda λ	Mu μ
Pango lineages	B	B.1.1.7	B.1.351 B1.351.2 B1.351.3	P.1 P.1.1 P.1.2	B.1.617.2 AY.1 AY.2 AY.3 AY.4 AY.5 AY.6 AY.7 AY.8 AY.9 AY.10 AY.11 AY.12	B.1.1.529 (BA.1, BA.2, BA.4, BA.5 BQ.1 XBB)	B.1.427/ B.1.429	P.2	B.1.525	P.3	B.1.526	B.1.617.1	C.37	B.1.621
GISAID clade/ lineages	S	GRY formerly GR/ 501Y.V1	GH/501Y.V2	GR/ 501Y.V3	G/452R.V3	GR/484A	GH/452R.V1	GR	G/484 K.V3	GR	GH	G/452R.V3	GR/ 452Q.V1	GH
Nextstrain clade	19A	20I	20H	20J	21A 21I 21J	21K	21C	20B	21D	21E	21F	21B	21G	21H
Earliest reported region	China	United Kingdom	South Africa	Brazil	India	South Africa	United States	Brazil	United Kingdom and Nigeria	Philippine	United States	India	Peru	Colombia
Variant of concern		√	√	√	√	√								
Variant of interest													√	√

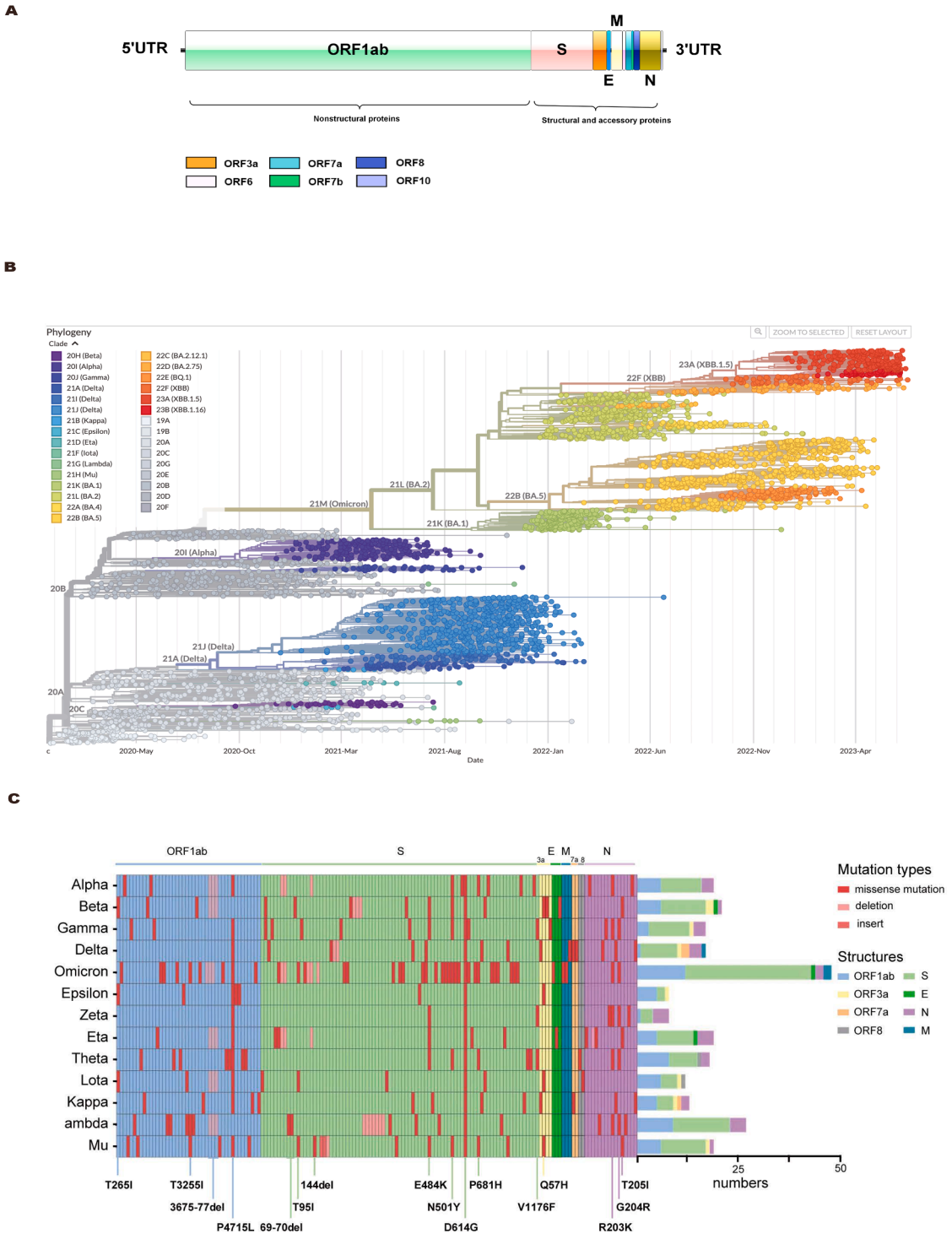

Fig. 1. Genome structure, phylogenetic analysis and mutation landscape of SARS-CoV-2 variants.
A. Genome structure of SARS-COV-2. (Refseq: NC_045512), Nonstructural proteins: ORF1ab, structural and accessory proteins spike glycoprotein (S), membrane (M), envelope protein (E), nucleocapsid (N), and accessory proteins of ORF3a, ORF6, ORF7, ORF8 and ORF10;
B. Phylogenetic tree of SARS-CoV-2 variants that were analyzed in the present study. Lengths of the tree branches represent relative phylogenetic distances between the clones and variants. Labels on the right denote the variants in the present study. Time axis denotes the time of SARS-CoV-2 variants exist.
C. Mutation landscape of SARS-CoV-2 variants. Mutation site and mutation type have been highlighted in different color as annotation in bottom right corner of picture.

structures were extracted from each trajectory between 0 and 50 ns and Molecular Mechanics-Poisson Boltzman Surface Area (MM-PBSA) was calculated using the gmx_mmpbsa.bsh script in gmxtools software.

2.5. *Pseudotype infection assay*

The pseudotype viruses were produced by co-transfection 293T cells with psPAX2, pLenti-GFP-luci, and plasmids encoding the coronavirus spike or VSV-G using Lipo8000 (Beyotime, China). The supernatants

were harvested at 72 h and 96 h post transfection, passed through a 0.45 μm filter, and centrifuged at 3500 rpm for 30 min to remove cell debris. Then the supernatants were stored at 4 °C overnight after the addition of PEG-6000 at a final concentration of 8.5 % and NaCl at a final concentration of 0.3 M. The mixtures were centrifuged at 3500 rpm for 30 min. Then, the viral pellets were resuspended by PBS, aliquoted, and stored at −80 °C for later use. The titer of pseudotype viruses was determined by p24 ELISA kit (Biodragon, China). For the pseudotype virus infection assay, the cells were seeded at 2×10^4 cells per well in a 96-well plate for 1 day and then inoculated with 10 μL media containing the pseudotype viruses (Multiplicity of Infection, MOI = 2.5) overnight. Then the medium was changed with fresh medium and cells were incubated for another 48 h followed by luciferase activity assay using the ONE-Glo™ Luciferase Assay System/Luciferase Assay System (Promega, USA) by a GloMax luminometer (Promega, USA).

3. Results

3.1. Sequence data selection and phylogenetic analysis

The protein annotation by Wuhan-Hu-1 (NCBI: NC_045512) was shown in Fig. 1A. We used the Nextstrain website to visualize the evolution of SARS-CoV-2. As shown in Fig. 1B, the Alpha variant was first reported globally in May 2020, followed by the emergence of the Beta, Gamma, and Epsilon variants, subsequent to the Delta mutant, discovered in May 2021, which became the globally prevalent strains until the emergence of Omicron in December 2021 that remains the predominantly prevalent strain to date (April 2023). To conduct an evolutionary analysis of these variants, a total of 3216 sequences representing 13 prominent SARS-CoV-2 variants worldwide were obtained from the GISAID database. Evolutionary tree analysis was performed, revealing that these 13 mutants were primarily segregated into four distinct clusters. Despite the dispersed distribution of most mutants, the Beta, Iota, Mu, and Epsilon mutants tended to cluster together. On the other hand, the Omicron, Gamma, Lambda, Alpha, and Eta mutants were found to be closely associated with each other. In contrast, the Delta variant appeared to display a greater degree of independence compared to other variants.

3.2. Sequence variations on structural proteins of SARS-CoV-2

To further study mutations of the variants, thirteen major SARS-CoV-2 variants globally involving 130 amino acids sequence retrieved stochastically from NCBI were aligned with Wuhan-Hu-1 (NCBI: NC_045512), respectively in this study. All common mutation sites within one variant were shown in Table 2 and Fig. 1C. A total of 159 mutations were found, including missense mutations, deletions, and insertions. Most of them were found in the S protein ($n = 85$). we found several common mutations in the S proteins of the variants. For instance, D614G in S protein existed in all variants. E484K in RBD domain of S protein existed in Beta, Gamma, Zeta, Eta, Theta, Mu, and similar E484A existed in Omicron. N501Y in the RBD domain existed in Alpha, Beta, Gamma, Omicron, Theta and Mu. T95I in the NTD existed in Omicron, Iota, and Mu. 69–70del in the NTD existed in Alpha, Omicron and Eta. V1176F in the CD domain existed in Gamma, Zeta and Theta. Mutation of amino acid at position H655Y in S protein existed in Gamma and Omicron. We also found other similar mutations. For instance, L452R mutation in Delta whereas L452Q in Lambda. P681H in Alpha, Omicron, Mu and Theta whereas P681R in Kappa and Delta, and K417N or K417T mutation in Beta and Gamma. Among 13 major variants, Omicron obviously has the most abundant mutation sites ($n = 29$) whereas Epsilon has the least mutation sites ($n = 2$). The over-abundant mutation sites might lead to the potential of Omicron for expansion and replacement of prior variants, for continually causing new waves with increased circulation in COVID-19 pandemic.

Some of the common mutations with amino acid substitutions were present in the N protein ($n = 16$). For instance, E203K existed in Alpha, Gamma, Omicron, Zeta, Theta and Lambda, whereas a similar mutation T205I existed in Beta, Eta, Mu. In terms of mutational changes, membrane (M) and envelope proteins (E) are relatively more conserved than N and S proteins. Accessory proteins ORF3a ($n = 4$), 7a ($n = 2$), and 8 ($n = 2$) also existed stable mutation. Chu DKW et al., reported that ORF3a-Q57H mutation caused the truncation of ORF3b, which led to the virus evasion of induction of cytokine, chemokine, and interferon-stimulated gene expression in primary human respiratory cells (Chu et al., 2021). Therefore, ORF3a with Q57H in Beta, Epsilon, Iota, Mu and ORF3a with S26L in Delta, Kappa may influence virus entry. In addition, Su et al. reported that ORF7a and ORF7b were associated with localization in the Golgi compartment and promote antagonism of interferon response after its ubiquitination (Cao et al., 2021). Thus, ORF7a with V82A in Delta and Kappa may also have an impact on the viral infectivity. These results suggest that mutation landscapes usually revealed similar tendencies with their evolutionary patterns (Young et al., 2020).

3.3. Sequence variations on non-structural proteins of SARS-CoV-2

We revealed 44 mutations in the nonstructural proteins of ORF1ab. Some mutations have a critical influence on SARS-CoV-2 function. For instance, T265I mutation located in the N-terminal domain of nsp2 of ORF1ab in Beta, Epsilon and Iota may affect the binding ability to prohibitin 1 and prohibitin 2 (PHB1 and PHB2) or localization at mitochondria-associated ER membranes (Cornillez-Ty et al., 2009; Davies et al., 2020; Lei et al., 2020b). In addition, we found that a 3675–3677 deletion in nsp6 (putative transmembrane domain) may influence the antagonizing type I interferon (IFN-I) response for increasing viral transmission and immune evasion (Xia et al., 2020). Moreover, P4715L mutation in the RNA-dependent RNA polymerase existed in all variants except Beta and Lambda may affect the severity and spread of the virus (Khailany et al., 2020; Badua et al., 2021). Among 13 major variants, Omicron has the most abundant mutation sites in ORF1ab ($n = 10$) while there is no any mutation site identified in other non-structural proteins in Omicron, Zeta, Eta, and Lambda. It is unclear whether this is mutation strategy with evolutionary advantage.

3.4. Comparison of physicochemical properties of S protein among 13 variants

Most of mutations were found in the S protein. Moreover, the S protein of SARS-CoV-2 plays a crucial role in viral entry into host cells and is a major target for vaccine development. Examination of the physicochemical properties (PCPs) of the S protein across variants may elucidate the evolutionary modifications that contributed to the enhanced infectivity of SARS-CoV-2. Previously, the potential roles of PCPs in SARS-CoV-2 proteins, encompassing molecular weight, thermal stability, and pH, have been investigated to establish quality control measures for vaccines (Shang et al., 2020). Through bioinformatics analyses, we assessed several physicochemical properties, including aliphatic index, grand average of hydropathicity, instability index, and isoelectric point (PI). Our analyses revealed notable differences in the physicochemical properties of the S protein among these variants. Variants such as Alpha, Omicron, Eta, Lambda and Mu, exhibited higher aliphatic compared to the wild type (WT) (Fig. 2A). Fig. 2B showed that Delta has the lowest grand average of hydropathicity, which may affect protein folding and stability. Additionally, instability index analysis identified specific residues with altered exposure, possibly influencing omicron's highly flexible structure to escape interactions with neutralizing antibodies (Fig. 2C). The PI values varied across the variants with Delta and Omicron being the two top highest value, indicating potential alterations in charge distribution (Fig. 2D). We then performed PCA analysis of the four physicochemical properties, the results showed that Omicron has a noticeable difference with others, followed by Delta and Theta (Fig. 2E).

Table 2
Mutational features of thirteen SARS-CoV-2 variants.

Gene	Variants of SASR-CoV-2												
	Alpha **α**	**Beta** **β**	**Gamma** **γ**	**Delta** **δ**	**Omicron** **o**	**Epsilon** **ε**	**Zeta** **ζ**	**Eta** **η**	**Theta** **θ**	**Iota** **ι**	**Kappa** **κ**	**Lambda** **λ**	**Mu** **μ**
S	69–70del 144del N501Y A570D D614G P681H T716I S982A D1118H	L18F D80A D215G 242–244del K417N E484K N501Y D614G A701V	L18F P26S D138Y R190S K417T E484K N501Y D614G H655Y V1176F	T19R E156G 157–158del L452R T478K D614G P681R D950N	A67V, 69–70del, T95I G142D 143–145del, 211del L212I G339D S373P S375F S477N T478K E484A Q493R G496S Q498R N501Y Y505H T547K D614G H655Y N679K P681H N764K D796Y N856K Q954H N969K L981F	D614G L462R	E484K D614G V1176F	Q52R A67V 69–70del 144del E484K D614G Q677H F888L	E484K N501Y D614G P681H E1092K H1101Y V1176F	L5F T95I D253G D614G	E484Q D614G P681R Q1071H	G75V T76I 246–252del D253N L452Q F490S D614G T859N	T95I Y144T 145–146 ins 147del R346K E484K N501Y D614G P681H D950N
ORF1ab	P4715L T1001I A1708D 3675–77del	T265I K1655N K3353R 3675–77del	S1188L K1795Q P4715L	P4715L	K856R 2083del L2084I A2710T T3255I P3395H 3674–76del I3758V P4715L I5967V	T265I S3159T P4715L P5378L D5584Y	P4715L	T2007I 3675–77del P4715L	D1554G S2625F D2980N D3861E L3930F P4715L L5604F A5692V	T265I L3201P 3675–77del P4715L	T1567I T3646A P4715L M5753I K6711R	T1246I P2287S F2387V L3201P T3255I G3278S 3675–77del	T1055A T1538I T3255I Q3729R P4715L P5743S
E	–	P71L	–	–	T9I	–	–	L21F	–	–	–	–	–
M	–	–	–	I82T	Q19E A63T	–	–	–	–	–	–	–	–
N	D3L R203K G204R/P S235F	T205I	P80R R203K G204R	D63G R103M D377Y	R203K G204R	–	A119S R203K G204R M234I	S2Y 3del A12G T205I	R203K G204R		R203M D377Y	P13L R203K G204R G214C	T205I
ORF3a	–	Q57H S171L	S253P	S26L	–	Q57H	–	–	–	Q57H	S26L	–	Q57H
ORF6	–	–	–	–	–	–	–	–	–	–	–	–	–
ORF7a	–	–	–	V82A T120I	–	–	–	–	–	–	V82A	–	–
ORF7b	–	–	–	–	–	–	–	–	–	–	–	–	–
ORF8	–	–	–	–	–	–	–	–	K2Q	T11I	–	–	–
ORF10	–	–	–	–	–	–	–	–	–	–	–	–	–

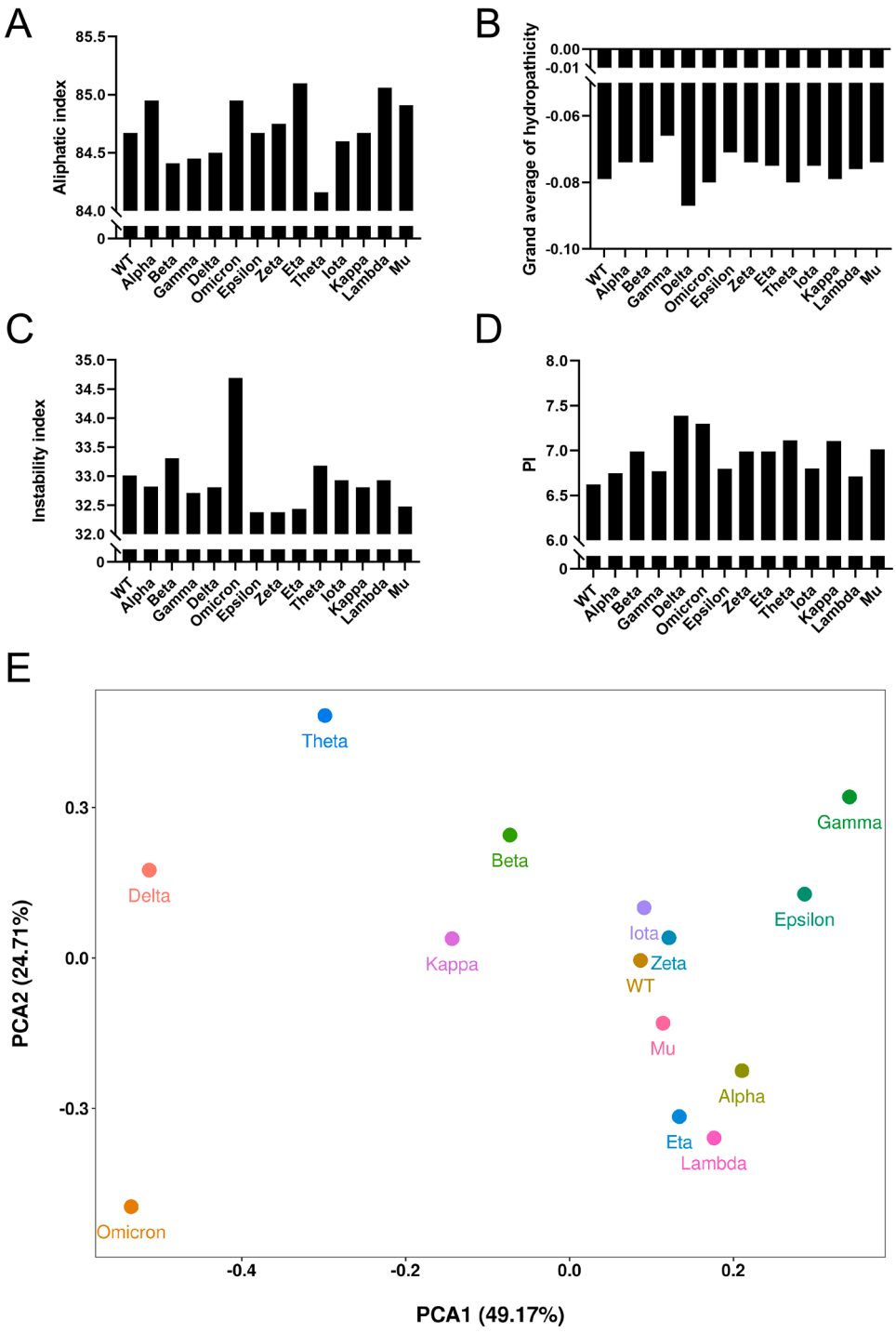

Fig. 2. Comparative analysis of physicochemical properties of the S protein among different variants.
The physicochemical property was analyzed via Expasy (https://www.expasy.org/). We uploaded the amino sequence of each variant spike protein and selected the different modules.
A. Column represents the aliphatic index of each variant;
B. Column represents the grand average of hydropathicity score of each variant;
C. Column represents the instability index of each variant;
D. Column represents the isoelectric point (PI) of each variant;
E. Based to the aliphatic index, Grand average of hydropathicity score, Instability index, and PI, the principal component analysis (PCA) was performed using omicstudio (https://www.omicstudio.cn/tool/144).

3.5. Comparative analysis of the mutation impact on the binding affinity between the RBD and ACE2

Considering the high mutation rate in the S protein and its significant role in the cell membrane fusion process (Table 3), we then examined the mutation impacts on receptor recognition and contextualized them in the protein structure. We first constructed the S protein structure of each variant and compared the RBD-ACE2 complex structure between

the WT and the variants. For the RBD structure, there were no obvious differences in the second, tertiary or quaternary structure. The α-helix, β-sheet, and β-turn were kept in all variants (Fig. 3A). Especially, Omicron with the highest mutation landscape still largely maintained its original structure of the wuhan-1. For instance, we found a mutation (L452R) at Delta RBD that mutated from neutral residue Leucine to positively charged Arginine residue. We also found E484 at Delta, Omicron and Mu RBD that mutated from acidic amino acid residue to basic charged lysine residue (K) or nonpolar aliphatic Alanine (A) residue. K417 at Beta and Gamma RBD that mutated from positively charged Arginine residue to neutral residue Asparagine.

Molecular dynamics simulations are computational modelling techniques used to study the molecular-scale behavior of materials (Fig. 3B–3E). To further explore whether these minor changes in the mutated sites of the RBD would affect the interaction between RBD and ACE2, Omicron, Delta and Beta were selected as representatives for performing molecular dynamics simulation due to their representative amino acid mutations and their neutralization significance. As shown in Fig. 3B, Omicron reached a low Root Mean Square Deviation (RMSD) in the early 20 ns while the WT had higher RMSF values (Fig. 3C) at the receptor binding motif (RBM, residues 370–380). In addition, the number of hydrogen bonds is also a key factor in the stability of the secondary and tertiary structure of proteins. The intramolecular hydrogen bond plots showed that most hydrogen bonds were observed in the Omicron RBD-ACE2 complex, followed by the Delta counterpart, while WT and Beta RBD-ACE2 complexes had less hydrogen bonds, indicating the stronger binding affinity of the RBD to ACE2 in Omicron and Delta (Fig. 3D). We also used the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method to calculate the binding free energy of each RBD-ACE2 complex. Consistent with the above results, the Omicron RBD-ACE2 complex with the largest binding free energy showed the most stable dynamic, suggesting that substitution mutations in the Omicron RBD may increase the stability of the RBD-ACE2 complex, and Delta was close to the Omicron (Fig. 3E).

The 493 sites of the S protein in WT exhibits hydrogen bonding interactions, which were also found at the same site as Delta and Omicron, but with different mutation types (Fig. 4). Table 4 shows the residual energy contribution of mutant residues compared to all three wild-type complexes. According to the results, this site contributes significantly to the binding affinity. However, it was not observed in the Beta variant, which may help explain its rapid disappearance. Additionally, Delta and Omicron variants exhibited increased intrachain interactions within the S protein, suggesting that the stabilization of the S protein, increased binding energy and intra-chain interactions are crucial. Moreover, the Delta variant achieved higher transmission through increased S protein stabilization and intra-chain interactions. In the Omicron S protein, the enhanced hydrogen-bond interaction and binding affinity with ACE2 could explain the higher transmissivity of this variant.

3.6. Immune escape analysis of SARS-CoV-2 variants

SARS-CoV-2 may develop immune escape mutations that could diminish the effectiveness of vaccines or immune responses resulting from previous infections, thus evading recognition and attack by the host immune system. Comparative analysis of the SARS-CoV-2 sequence and evolution has identified potential immune escape mutations, but further research is needed to understand their actual impact on the immune escape capabilities of different strains. The deep mutation scanning method based on the yeast system provides sufficient immune escape data. We calculated the immune escape scores of all sites on the S protein of 13 strains based on this data. As shown in Fig. 5, we found 10 immune escape sites among 13 major variants, some of which are consistent with previous reports. Mutation site 505 has the highest immune escape score, indicating the potential importance of mutations at this site for the escape capability of the strain. Additionally, the Omicron variant exhibits more immune escape mutation sites compared to other variants, such as 493, 496, 498, and 501. Site 484 is conserved in all variants, posing a shared challenge for antibody screening and design, as mutations at this site have been associated with escape from human polyclonal serum antibodies. Notably, four immune escape mutation sites (339/337/477/496) have not been reported yet and are specific to Omicron, whereas 462 is specific to Epslion. These results provide important information for predicting the infectivity of new variants, designing vaccines, and developing therapeutic antibody.

3.7. Experimental validation of the molecular dynamics simulation results using pseudotype virus infection assay

The results from molecular dynamics (MD) simulations analysis suggest a stronger RBD-ACE2 binding capacity of Omicron and Delta compared with WT and other variants, which was also consistent with their dominant positions during the pandemic period. To validate the simulation results, we generated pseudotype viruses bearing the S proteins of the WT, Omicron, Beta, and Delta, respectively, and examined

Table 3
Common mutations in the S proteins among thirteen SARS-CoV-2 variants.

Mutations	Variants	Location	Impact
D614G	all variants	near the cleavage site	Associated with higher viral load [30]
E484K(A)	Beta, Gamma, Zeta, Eta, Theta, Mu(Omicron)	RBM	Reducing neutralization of antibodies produced by previous infection or vaccination, enhanced ACE2 recognition, thus increased the viral infectivity and antigenicity [31, 32]
N501Y	Alpha, Beta, Gamma, Omicron, Theta, Mu	RBM	Increasing binding ability to the receptor and transmissibility than wild type [33, 34]
P681H®	Alpha, Omicron, Mu, Theta (Kappa, Delta)	near the cleavage site	Enhancing variant replication via increased S1/S2 cleavage and increasing infection via cell surface entry [40, 41].
T478K	Delta, Omicron	RBM	Critical for transmission [83, 88].
L452R (Q)	Delta(Lambda)	RBM	Promoting the interaction between spike and ACE2 and help to evade vaccine-stimulated antibodies [39] [83–85]; Promoting viral replication by increasing spike stability, infection and fusion, increase viral load and avoid the attack of CD8 T cells responsible for virus elimination [86, 87].
K417N (T)	Beta(Gamma)	RBD	Creating new inter-protein contacts and changes the internal structural dynamics thereby increases the binding and eventually the infectivity [42, 43]
T95I	Omicron, Iota, Mu	NTD	Related to subsequent infection after successful vaccination [35].
69–70del	Alpha, Omicron, Eta	NTD	Impact on virus response to inactivated-virus vaccine [36]
V1176F	Gamma, Zeta, Theta	CD domain	Affecting the viral pathogenesis through the alteration of protein conformation, leading to a difference in transmission and virulence [37]
H655Y	Gamma, Omicron	near the cleavage site	Escaping from human monoclonal antibodies [38].

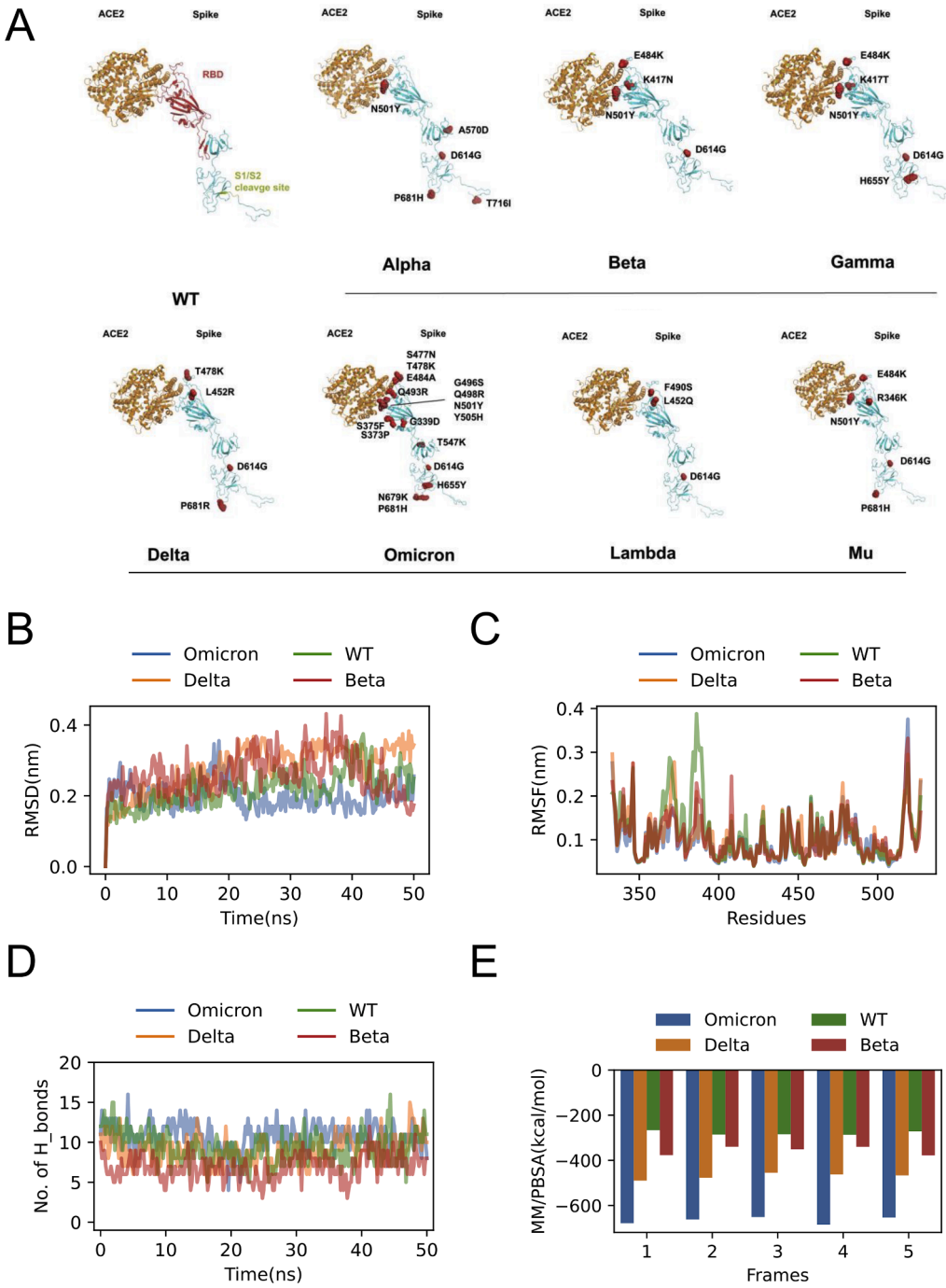

Fig. 3. Molecular dynamics analysis of SARS-CoV-2 S protein (AA 319–720) complexed with human ACE2 of the Mutated residues compared with the wildtype for all the thirteen complexes.

A. The hot pink-colored sphere shape residues indicate the mutation in the S protein of SARS-CoV-2;

B. The RMSD plot demonstrates a consistent deviation among all simulated structures. C. The RMSF plot indicates increased flexibility in Residues 350–400 of the RBD, whereas the mutated residues show reduced fluctuation and are similar across all three structures.

D. The hydrogen bond count reveals that both the wildtype and Kappa variant have a similar and higher number of hydrogen bonds compared to the Delta and Alpha variants.

E. The MM/GBSA binding free energy of the 20 structure complexes extracted from each trajectory, over an equal span, suggests that both Kappa and Alpha S variants exhibit higher affinity for hACE2 compared to Delta and wildtype.

the infectivity of each of the pseudotype viruses in 293T and 293T/hACE2 cells using the luciferase activity assay. As expected, these cells were more susceptible to infections by the Omicron and Delta pseudotype viruses, validating the MD simulation results (Fig. 6).

4. Discussion

In this study, we analyzed the genome sequences of 13 major variants of SARS-CoV-2 and revealed 159 stable mutations. Over half of these mutations (53 %) occurred in the S protein of SARS-CoV-2. Previous studies have shown that the S protein is essential in the virus entry

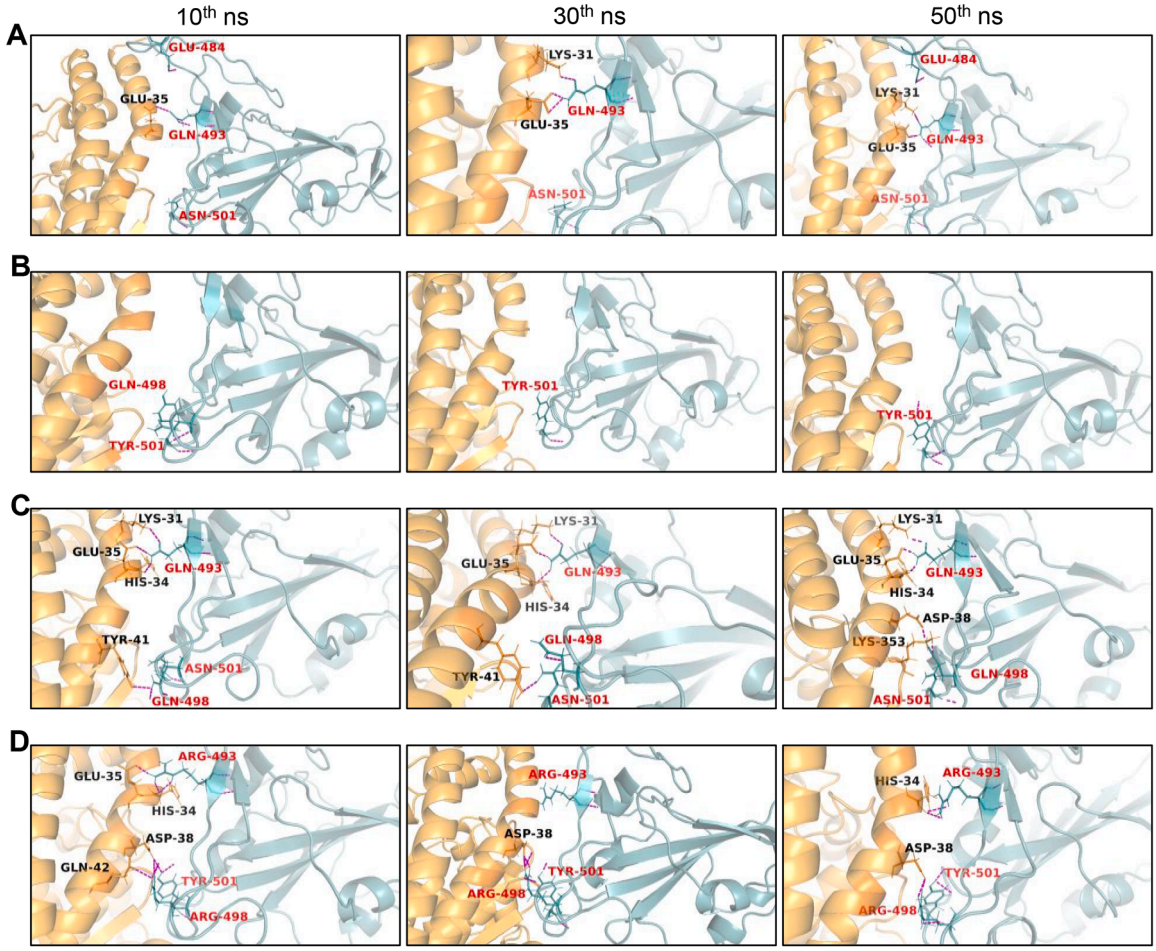

Fig. 4. Schematic diagrams of molecular dynamics analysis of SARS-CoV-2 RBD complexed with hACE2 at 10, 30, 50 ns.
A. The interactions of wildtype residues at the 50th, 100th, and 150th ns of the simulation revealed that GLU-35 of the hACE2 has a single hydrogen bond interaction with GLN-493 of the S protein. Similarly, LYS-31 of the hACE2 protein only interacts through hydrogen bonds with GLN-493.
B. TYR-501 of the S protein only interacts through hydrogen bonds with its own residues.
C. An additional hydrogen bond between HIS-34 and ARG-493 was observed in all three frames of the S protein of Delta variant.
D. Interactions of the S protein of the Omicron variant include a hydrogen bond between TYR-501 of the mutant S protein and ASP-38 of hACE2.

Table 4
Residue wise energy contribution of the mutated residues compared with the wildtype for all the three complexes.

WT (kcal/mol)						Beta (kcal/mol)		Delta (kcal/mol)			Omicron (kcal/mol)			
L452	E484	T478	Q493	Q498	N501	K484	Y501	R452	K478	N501	K478	R493	R498	Y501
−0.39	52.46	−0.16	−1.04	1.63	−1.86	−56.30	−2.37	−50.98	−46.55	−2.69	−43.31	−60.97	−68.30	−4.14
−0.36	48.75	0.58	−2.64	1.04	−1.55	−51.25	−0.37	−49.69	−45.16	−1.47	−45.60	−65.28	−65.27	−4.48
−0.30	49.43	0.98	−1.06	0.37	−2.23	−55.03	−1.85	−46.87	−41.26	−2.89	−42.43	−66.18	−67.74	−3.97
−0.52	52.95	0.66	−3.45	−0.44	−2.42	−51.33	−3.64	−52.16	−45.76	−2.39	−45.89	−63.59	−71.70	−4.07
−0.24	51.44	−0.11	−2.68	1.80	−2.96	−53.51	−3.42	−49.93	−43.60	−4.48	−39.29	−66.07	−69.29	−4.53
−0.36	**51.01**	**0.39**	**−2.17**	**0.88**	**−2.20**	**−53.48**	**−2.33**	**−49.93**	**−44.47**	**−2.78**	**−43.30**	**−64.42**	**−68.46**	**−4.24**
±0.09	**±1.65**	**±0.45**	**±0.96**	**±0.83**	**±0.48**	**±2.00**	**±1.18**	**±1.76**	**±1.87**	**±0.98**	**±2.40**	**±1.96**	**±2.09**	**±0.23**

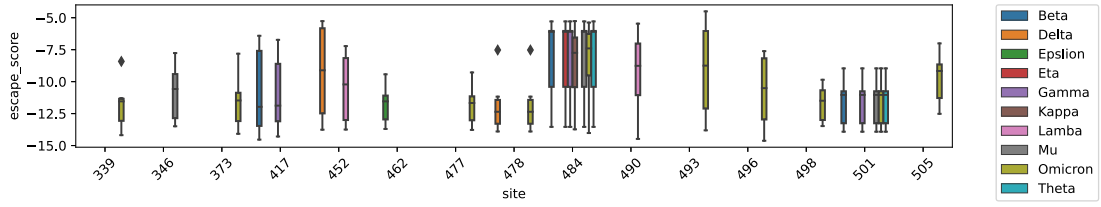

Fig. 5. Comparative immune escape ability.
The immune escape scores of different variants were calculated based on the deep mutational scanning data from https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS_Omicron, and different colored boxplots presented the corresponding variants' immune escape scores.

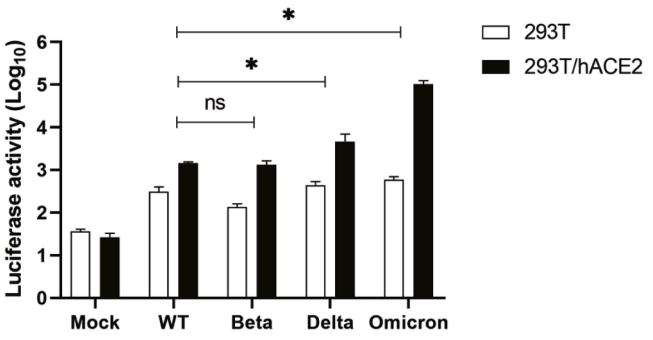

Fig. 6. The infectivity of pseudotype viruses in the susceptible cell lines. Entry of SARS-CoV-2 S pseudotyped virions into human cell lines 293T (white) and 293T/hACE2 (black). Cells were infected with each SARS-CoV-2 variant pseudotyped viruses, respectively. At 72 h post infection, viral entry efficiency was measured by luciferase activity assay. Significant difference from mock were determined by two-tailed unpaired *t*-test. $^{*}P < 0.05$. Error bars indicate SD ($n = 3$).

into host cells (Lei et al., 2020a; Walls et al., 2020). Numerous mutation sites in the S protein were associated with the gradual adaptation of SARS-CoV-2 to human hosts and the reduced efficacy of current drugs and vaccines (van Dorp et al., 2020). Therefore, additional comparative analysis of the relationship between mutation sites and their functions is crucial for understanding the pathogenicity, progression, prevention, and treatment of the disease caused by SARS-CoV-2.

A number of mutations on the S protein, particularly those containing positively charged amino acids, not only influence the interaction with ACE2, but also boost the binding ability to receptors on cellular surfaces with a negative charge, intramolecular interactions, and the conformational stability of the S protein. By comparing and analyzing the physicochemical properties of the S protein (Fig. 2), we found that compared to the wild type (WT), variants such as Alpha, Omicron, Eta, Lambda, and Mu demonstrated higher indices of hydrophobicity (Fig. 2A). In Fig. 2B, Delta had the lowest average hydrophobicity, which may affect the folding and stability of proteins. Additionally, the analysis of the instability index identified specific amino acid residues with altered exposure, potentially impacting the structure with high flexibility of Omicron and its interaction with the neutralization antibodies (Fig. 2C). The average hydropathy values (GRAVY) varied among different variants, indicating possible changes in the distribution of charges, with Delta and Omicron having the highest values (Fig. 2D). This provides insight into the molecular basis of the increased infectivity, transmission, and immune evasion of Omicron and Delta variants. Bioinformatic tools are powerful techniques for understanding the physical and chemical properties, structure, function, and potential epitopes of the S protein, thus providing a reference for the research and prevention of novel coronavirus infection. Comparative analysis of the physical and chemical properties of the S protein also provides valuable information regarding the potential functional and structural differences of the S protein.

The Omicron variant harbors almost 30 mutations in the S protein, 15 of them located in the RBD region. Molecular dynamics simulations indicate a potential correlation between these mutations and an increased binding affinity to the ACE2 receptor. Consistent with previous studies (Yi et al., 2020), simulation results demonstrate that the mutated residues Q493R and Q498R interact with ACE2 through the formation of multiple hydrogen bonds, thereby enhancing the overall stability of the Omicron S protein. Based on these unique mutations, we reasoned that the recently emerged Omicron variants employs a more efficient strategy, potentially influencing the potency of current drugs and vaccines, and subsequently escalating the transmission rates. Among the mutations in the RBD region, two specific mutation sites, namely L452R and T478K, significantly augmented the binding affinity to ACE2, and played critical roles in immune escape and transmission, which were consistent with the previous reports by Jhun et al. and Andreano et al. (2021).

The Molecular dynamics analysis of the SARS-CoV-2 S protein complexed with ACE2 receptor unveils that, in comparison to the wild type and other variants, Omicron and Delta exhibited an increased number of hydrogen bonds and higher binding free energy within the complex. This enhanced binding capability between the RBD and ACE2 and the ensuing stable conformation is further supported by pseudotype virus infection assay. Integrating the protein mutation patterns observed in Omicron and Delta with the results of molecular dynamics analysis may potentially provide significant insights into elucidating their prominent roles in the ongoing pandemic.

Additionally, we calculated the immune escape scores for various strains at distinct sites. In comparison to other variants, the Omicron variant exhibits a greater number of mutation sites associated with immune escape. Specifically, a total of 10 immune escape sites, including 493, 496, 498, and 501, are identified within the RBD region. Notably, site 505 attains the highest immune escape score, highlighting the potential importance of mutations at this site in enabling the strain to evade the immune response. Furthermore, site 484 is consistently conserved across all variants. As reported, these two mutations both might lead to a more significant escape from immune protection elicited by previous SARS-CoV-2 infection and serum antibodies, presenting a shared challenge in antibody screening and design (Kumar et al., 2022). By employing rigorous strain selection and employing diverse analytical methods, we acquired comparative analysis findings regarding immune escape across different strain S proteins. This endeavor yields valuable insights for future research and vaccine development. In addition, we observed a similarity in binding affinity between Omicron RBD-ACE2 and WT RBD-ACE2, suggesting that compensatory mutations may account for immune evasion and transmission capabilities. Consequently, a comprehensive understanding of these variations could aid in the development of targeted therapies and vaccine strategies to address viral evolutionary traits.

In conclusion, we conducted a systematic analysis of sequence and structural alignments, molecular dynamics, and immune escape scoring for 13 major SARS-CoV-2 variants, with a focus on the dominant Omicron and Delta variants. Viral mutations play a crucial role in the pathogenicity of SASR-CoV-2, potentially leading to antigenic changes, tissue targeting diversity, and weakened host immune responses (Wan et al., 2020; Giordano et al., 2021). In particular, we have identified shared mutations within the S protein of Omicron. Our findings may not only provide novel insights into understanding the impact of various mutations on the pathogenicity of SARS-CoV-2, but may also help develop novel strategies against the ongoing infections of SARS-CoV-2.

CRediT authorship contribution statement

Han Bai: Writing – original draft, Methodology, Formal analysis, Data curation. **Xuan Zhang:** Writing – original draft, Software, Methodology. **Tian Gong:** Writing – original draft, Methodology. **Junpeng Ma:** Methodology. **Peng Zhang:** Methodology. **Zeqiong Cai:** Methodology. **Doudou Ren:** Methodology. **Chengsheng Zhang:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Conceptualization, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We are very grateful to other team members from our laboratories at The First Affiliated Hospital of Xi'an Jiaotong University and The First Affiliated Hospital of Nanchang University. This study is supported in part by the operational funds from The First Affiliated Hospital of Xi'an Jiaotong University and The First Affiliated Hospital of Nanchang University (CZ).

References

Andreano, E., et al., 2021. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. Proc. Natl. Acad. Sci. USA 118 (36).
Arunachalam, P.S., et al., 2021. Adjuvanting a subunit COVID-19 vaccine to induce protective immunity. Nature 594 (7862), 253–258.
Badua, C., et al., 2021. Genomic and proteomic mutation landscapes of SARS-CoV-2. J. Med. Virol. 93 (3), 1702–1721.
Cai, Y., et al., 2020. Distinct conformational states of SARS-CoV-2 spike protein. Science 369 (6511), 1586–1592.
Campbell, F., et al., 2021. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill. 26 (24).
Cao, Z., et al., 2021. Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. Cell. Mol. Immunol. 18 (3), 746–748.
Chakraborty, C., et al., 2021. Evolution, mode of transmission, and mutational landscape of newly emerging SARS-CoV-2 variants. mBio 12 (4), e0114021.
Chu, D.K.W., et al., 2021. Introduction of ORF3a-Q57H SARS-CoV-2 variant causing fourth epidemic wave of COVID-19, Hong Kong, China. Emerg. Infect. Dis. 27 (5), 1492–1495.
Cornillez-Ty, C.T., et al., 2009. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. J. Virol. 83 (19), 10314–10318.
Cosar, B., et al., 2022. SARS-CoV-2 mutations and their viral variants. Cytokine Growth Factor Rev. 63, 10–22.
Davies, J.P., et al., 2020. Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus nonstructural proteins identifies unique and shared host-cell dependencies. ACS Infect. Dis. 6 (12), 3174–3189.
de Souza, A.S., et al., 2022. Molecular dynamics simulations of the spike trimeric ectodomain of the SARS-CoV-2 Omicron variant: structural relationships with infectivity, evasion to immune system and transmissibility. J. Biomol. Struct. Dyn. 41 (19), 9226–9343.
Desingu, P.A., Nagarajan, K., 2022. The emergence of Omicron lineages BA.4 and BA.5, and the global spreading trend. J. Med. Virol. 94 (11), 5077–5079.
Foster, P.L., 2000. Adaptive mutation: implications for evolution. Bioessays 22 (12), 1067–1074.
Giordano, D., et al., 2021. Structural dissection of viral spike-protein binding of SARS-CoV-2 and SARS-CoV-1 to the human angiotensin-converting enzyme 2 (ACE2) as cellular receptor. Biomedicines 9 (8).
Goyal, A., et al., 2021. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. eLife 10.
Greaney, A.J., et al., 2021. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host Microbe 29 (1), 44–57 e49.
Hadfield, J., et al., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34 (23), 4121–4123.
Hatcher, E.L., et al., 2017. Virus variation resource - improved response to emergent viral outbreaks. Nucleic Acids Res. 45 (D1), D482–d490.
Hung, J.H., Weng, Z., 2016. Sequence alignment and homology search with BLAST and ClustalW. Cold Spring Harb. Protoc. 2016 (11).
Izumi, H., et al., 2020. SSSCPreds: deep neural network-based software for the prediction of conformational variability and application to SARS-CoV-2. ACS Omega 5 (47), 30556–30567.
Khailany, R.A., et al., 2020. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 19, 100682.
Kirchdoerfer, R.N., et al., 2018. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. Sci. Rep. 8 (1), 15701.
Kistler, K.E., et al., 2022. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. Cell Host Microbe 30 (4), 545–555 e544.
Kumar, S., et al., 2022. Omicron (BA.1) and sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: a comparative sequence and structural-based computational assessment. J. Med. Virol. 94 (10), 4780–4791.
Lan, J., et al., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581 (7807), 215–220.
Lei, C., et al., 2020a. Neutralization of SARS-CoV-2 spike pseudotyped virus by recombinant ACE2-Ig. Nat. Commun. 11 (1), 2070.
Lei, X., et al., 2020b. Activation and evasion of type I interferon responses by SARS-CoV-2. Nat. Commun. 11 (1), 3810.
Li, Q., et al., 2020. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 182 (5), 1284–1294.e1289.
Lill, M.A., Danielson, M.L., 2011. Computer-aided drug design platform using PyMOL. J. Comput. Aided Mol. Des. 25 (1), 13–19.
Mannar, D., et al., 2022. SARS-CoV-2 Omicron variant: antibody evasion and cryo-EM structure of spike protein-ACE2 complex. Science 375 (6582), 760–764.
Metzgar, D., Wills, C., 2000. Evidence for the adaptive evolution of mutation rates. Cell 101 (6), 581–584.
Mlcochova, P., et al., 2021. SARS-CoV-2 B1.617.2 Delta variant replication and immune evasion. Nature 599 (7883), 114–119.
Muecksch, F., et al., 2021. Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations. Immunity 54 (8), 1853–1868 e1857.
Ponde, R.A.A., 2023. Physicochemical effects of emerging exchanges on the spike protein's RBM of the SARS-CoV-2 Omicron subvariants BA.1-BA.5 and its influence on the biological properties and attributes developed by these subvariants. Virology 587, 109850.
Qu, P., et al., 2023. Enhanced neutralization resistance of SARS-CoV-2 Omicron subvariants BQ.1, BQ.1.1, BA.4.6, BF.7, and BA.2.75.2. Cell Host Microbe 31 (1), 9–17 e13.
Rani, P.R., et al., 2021. Symptomatic reinfection of SARS-CoV-2 with spike protein variant N440K associated with immune escape. J. Med. Virol. 93 (7), 4163–4165.
Roy, A., et al., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5 (4), 725–738.
Sarkar, A., et al., 2023. The relative prevalence of the Omicron variant within SARS-CoV-2 infected cohorts in different countries: a systematic review. Hum. Vaccines Immunother. 19 (1), 2212568.
Shang, J., et al., 2020. Cell entry mechanisms of SARS-CoV-2. Proc. Natl. Acad. Sci. USA 117 (21), 11727–11734.
Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22 (13).
Telenti, A., et al., 2022. The Evolution and Biology of SARS-CoV-2 Variants. Cold Spring Harb. Perspect. Med. 12 (5).
Thorne, L.G., et al., 2022. Evolution of enhanced innate immune evasion by SARS-CoV-2. Nature 602 (7897), 487–495.
Van Der Spoel, D., et al., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26 (16), 1701–1718.
van Dorp, L., et al., 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect. Genet. Evol. 83, 104351.
Walls, A.C., et al., 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181 (2), 281–292.e286.
Wan, Y., et al., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J. Virol. 94 (7).
Wang, C., et al., 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. J. Med. Virol. 92 (6), 667–674.
Waterhouse, A., et al., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46 (W1), W296–w303.
(01 June 2020). "WHO Coronavirus (COVID-19) Dashboard." from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.
"World Health Organization. Weekly Epidemiological Update on COVID-19: 11 May 2021. ." from https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19-11-may-2021.
Wu, F., et al., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579 (7798), 265–269.
Xia, H., et al., 2020. Evasion of Type I interferon by SARS-CoV-2. Cell Rep. 33 (1), 108234.
Xia, S., et al., 2022. Origin, virological features, immune evasion and intervention of SARS-CoV-2 Omicron sublineages. Signal Transduct. Target. Ther. 7 (1), 241.
Yadav, R., et al., 2021. Role of structural and non-structural proteins and therapeutic targets of SARS-CoV-2 for COVID-19. Cells 10 (4).
Yi, C., et al., 2020. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. Cell. Mol. Immunol. 17 (6), 621–630.
Yoshimoto, F.K., 2020. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. Protein J. 39 (3), 198–216.
Young, B.E., et al., 2020. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet 396 (10251), 603–611.
Zhou, B., et al., 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. Nature 592 (7852), 122–127.