ACCSCIENCE PUBLISHING

Microbes & Immunity

ORIGINAL RESEARCH ARTICLE

A proposal for biologically relevant classification of SARS-CoV-2 variants

Saja Al-Baidhani[1], Tarneem Sabra[1], Aslam Al-Baidhani[1], Mohammed Sallam[2,3,4], and Malik Sallam[1,5*]

[1]Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman, Jordan

[2]Department of Pharmacy, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai, United Arab Emirates

[3]Department of Management, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai, United Arab Emirates

[4]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

[5]Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman, Jordan

Abstract

The present classification of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants plays a central role in shaping public health policies, vaccine strategies, and global risk communication. However, existing designations of variants of concern (VOCs) rely on evolving epidemiological and phenotypic criteria rather than quantitative genetic divergence thresholds. In this study, we evaluated the genetic divergence of SARS-CoV-2 variants relative to human immunodeficiency virus type 1 (HIV-1), hepatitis C virus (HCV), and influenza A virus, and proposed a framework integrating genetic, functional, and epidemiological criteria for variant classification. Comparative phylogenetic analysis assessed the divergence of SARS-CoV-2 (*S*) relative to HIV-1 (*env*), HCV (*E1*), and influenza A virus (*HA*). Maximum likelihood phylogenies with bootstrap support were constructed using MEGA6, and pairwise genetic distances were calculated through the maximum composite likelihood model. Monte Carlo simulations (n = 1,000) using adjusted SARS-CoV-2 evolutionary rates (0.0006 – 0.003 substitutions/site/year) estimated time to reach divergence thresholds defined by other viruses. SARS-CoV-2 variants showed a maximum divergence of 0.006 substitutions/site – far below thresholds for HIV-1 (0.157), HCV (0.371), and influenza A (1.956). Projections estimate HIV-1-like divergence in 53.7 years, HCV-like in 126.8 years, and influenza A-like in 668.6 years. No present VOC met all proposed functional criteria: transmissibility, immune escape, disease severity, and global dominance. Omicron exhibited partial immune escape but insufficient divergence for lineage reclassification. While present classification supports short-term response, integrating evolutionary benchmarks would enhance their biological relevance as the virus continues to diversify. A new evidence-based framework is needed to reduce public alarm, guide rational policymaking, and prioritize durable countermeasures over variant-specific responses.

Keywords: Phylogeny; Classification; COVID-19; Lineage; Variant

***Corresponding author:**
Malik Sallam
(malik.sallam@ju.edu.jo)

Citation: Al-Baidhani S, Sabra T, Al-Baidhani A, Sallam M, Sallam M. A proposal for biologically relevant classification of SARS-CoV-2 variants. *Microbes & Immunity*. 2025;2(3):87-107.
doi: 10.36922/MI025190042

Received: May 10, 2025

1st revised: May 30, 2025

2nd revised: June 3, 2025

Accepted: June 4, 2025

Published online: June 17, 2025

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a member of the family *Coronaviridae* that was first characterized in early 2020.[1-3] This virus is the causative agent of coronavirus disease 2019 (COVID-19) and long COVID.[4,5] COVID-19 has been recognized as the pandemic of the 21st century.[6] It ranges in manifestations from asymptomatic infections and mild cold-like illnesses to more severe forms that result in significant morbidity and mortality, especially in elderly patients with chronic illnesses.[7-11]

Like other members of *Coronaviridae*, SARS-CoV-2 has a genome comprising a positive-sense, single-stranded RNA with a linear configuration.[12,13] Despite the lack of accurate characterization of the exact origins of SARS-CoV-2, the early spread of the virus has been well documented.[14-18] Phylogenetic analyses of SARS-CoV-2 sequences from the early infections in China suggested that the first human cases likely occurred in late 2019.[19,20] The present evidence suggests that the emergence of SARS-CoV-2 resulted from recombination events between bat SARS-like coronaviruses and, possibly, coronaviruses found in pangolins, leading to cross-species transmission.[21-24] More recent research suggested that direct spillover from bats remains the most likely scenario, with limited genomic evidence supporting a significant role of pangolins in the emergence of SARS-CoV-2.[25,26]

The earliest available SARS-CoV-2 genomes were obtained from patients in December 2019.[27] These early viral sequences provided key insights into the genetic structure and early divergence of SARS-CoV-2.[3] Chinese researchers conducted phylogenetic analyses comparing these genomes with known bat and pangolin coronaviruses to infer the likely ancestral form of the virus.[28] Based on mutational differences, they identified an ancestral genome type, designated "S", and a more dominant derived type, labeled "L", to reflect specific amino acid substitutions.[29] Concurrently, Western researchers conducted independent analyses using similar methodologies but assigned different nomenclatures to describe the early divergence of SARS-CoV-2. Instead of "S" and "L," they labeled the ancestral strain as "A" and the derived strain as "B".[30-33] Over time, the B lineage continued to mutate, giving rise to sub-lineages, including B.1, which became the predominant SARS-CoV-2 lineage globally.[34] This lineage eventually diversified into major global variants of concern (VOCs), which the World Health Organization (WHO) formally labeled as Alpha, Beta, Gamma, Delta, and Omicron variants.[35] The early genomic studies illustrated the complex evolutionary dynamics of SARS-CoV-2, highlighting both early genetic bottlenecks and the subsequent rapid adaptation of the virus.[36]

During the initial phase of the COVID-19 pandemic, the relatively low number of infections limited the opportunities for genetic diversification within the SARS-CoV-2 genome. With fewer viral replication cycles, the accumulation of mutations was constrained, resulting in a lower frequency of SARS-CoV-2 variant emergence.[37] At this stage, mutations in the spike (S) protein, particularly within the receptor-binding domain (RBD), which mediates interaction with the human angiotensin-converting enzyme-2 (ACE2) receptor, were rarely observed.[37] As transmission increased globally, the virus evolved, with a positive selection of certain substitutions that conferred higher transmissibility (e.g., D614G).[38,39] Over time, specific S protein substitutions that enhanced SARS-CoV-2's ability to bind more efficiently to human cells became more prevalent and led to the emergence of genetic variants with epidemiological advantages. Notably, the Alpha (B.1.1.7) variant, first identified in the UK, and the Delta (B.1.617.2) variant, which originated in India, both demonstrated significantly higher transmission rates than earlier circulating strains.[40-42] These variants carried key S protein mutations, such as N501Y in the Alpha variant and L452R and P681R in the Delta variant, which enhanced their ability to infect host cells and possibly contributed to their global dominance at different phases during the pandemic.[43,44] The progressive emergence of more transmissible variants indicates the role of natural selection in shaping SARS-CoV-2 evolution, favoring mutations that increase infectivity while the virus continues to adapt to human hosts.[45]

At present, SARS-CoV-2 variants are designated as distinct viral lineages when they exhibit genetic changes deemed significant enough to warrant separate classification, as determined by the WHO Coronavirus Network (CoViNet).[35,46] Since the onset of the COVID-19 pandemic, the continuous evolution of SARS-CoV-2 has led to the identification of multiple VOCs and variants of interest (VOIs).[43,47,48] These designations have been based on the projected capacity of emerging variants to outcompete previously circulating strains through increased transmissibility or the necessity to adjust public health interventions.[49-51] This classification system has been important to facilitate epidemiological surveillance and risk assessment. However, its primary reliance on limited genetic divergence and evolving phenotypic criteria, rather than on consistent, functionally meaningful differentiation, highlights the opportunity for refining the system. Incorporating genetic and evolutionary benchmarks could enhance the alignment between nomenclature and the underlying virological or clinical significance of emerging SARS-CoV-2 variants.[52-54]

Certain SARS-CoV-2 variants are classified as VOCs due to their ability to maintain or enhance replication fitness despite increasing levels of population immunity – either through natural infection or vaccination.[43,45] These variants often exhibit mutations that provide a selective advantage, allowing them to spread more efficiently in partially immune populations.[55] A defining feature of many VOCs is the presence of mutations in the RBD of the S protein, which plays a critical role in host cell entry through the human ACE2 receptor.[56] Thus, the designation VOC is reserved for SARS-CoV-2 lineages in which specific genetic changes significantly enhance RBD binding affinity – as seen with substitutions such as N501Y – while also demonstrating epidemiological evidence of increased transmissibility.[57,58]

Before being categorized as a VOC, an emerging SARS-CoV-2 lineage is often first labeled as a VOI or, in some national surveillance frameworks, a variant under investigation.[35] If subsequent data confirm enhanced transmissibility, immune escape potential, or increased disease severity, the variant is formally designated as a VOC.[35] Once a variant reaches VOC status, it is systematically classified within the Pango nomenclature system, which provides detailed lineage assignments based on phylogenetic relationships.[48] In addition, the variant is assigned to specific clades within Nextstrain and Global Initiative on Sharing All Influenza Data (GISAID), two global platforms used for genomic epidemiology and viral evolution tracking.[59,60] These classification systems facilitate real-time monitoring of SARS-CoV-2 evolution, aiding in public health responses and vaccine adaptation strategies.

The WHO played a central role in the classification and monitoring of SARS-CoV-2 variants, regularly updating its framework to reflect emerging evidence on viral evolution, transmissibility, and immune escape potential.[35,47] Thus, the identification and designation of SARS-CoV-2 variants rely on genomic surveillance efforts, with submissions from member states analyzed through global platforms such as GISAID. This is followed by field investigations to assess the public health impact of SARS-CoV-2 variants.[61] Over time, the WHO expanded its classification system to include additional categories beyond VOCs. The updated framework now includes VOIs and variants under monitoring.[60] A VOI is defined as a variant possessing genetic changes that are predicted or known to affect viral characteristics, such as transmissibility, virulence, antibody evasion, therapeutic susceptibility, or detectability.[35,47] In addition, to be classified as a VOI, a variant must exhibit increasing circulation in at least one WHO region, raising concerns about potential global public health implications.[35] Other public health agencies, such as the United States Centers for Disease Control and Prevention (CDC), maintain independent classification criteria, which can lead to differences in the timing of variant de-escalation.[62,63] For example, the CDC, the European Centre for Disease Prevention and Control (ECDC), and the WHO de-escalated SARS-CoV-2 variants at different times, reflecting regional epidemiological assessments and risk evaluations.[35,64,65]

At this point, it is important to consider the foundational pillars in the field of virus evolution. It is well established that the accumulation of sufficient genetic divergence to produce a biologically distinct virus lineage – one with unique properties affecting transmission dynamics, pathogenicity, or immune interactions – typically requires several years of sustained evolutionary pressure.[66-70] The present classification framework for SARS-CoV-2, which designates new variants based on short-term genetic fluctuations, contradicts this fundamental understanding of viral evolution.[71] The issue is particularly pronounced in SARS-CoV-2 due to its genomic stability relative to other RNA viruses, such as human immunodeficiency virus type 1 (HIV-1), hepatitis C virus (HCV), and influenza A virus, all of which evolve under stronger selection pressures and exhibit markedly higher mutation rates.[72-75] Unlike HIV-1, HCV, and influenza A virus, SARS-CoV-2 possesses a proofreading exonuclease (nsp14-ExoN), a unique feature among the majority of RNA viruses that reduces the accumulation of replication errors.[76,77] This proofreading mechanism results in a significantly slower evolutionary rate for SARS-CoV-2, estimated between 0.0004 and 0.002 substitutions per site per year (s/s/y).[49,73,78] This evolutionary rate allows for gradual adaptation; however, it does not support the level of rapid divergence seen among other RNA viruses.[79] Consequently, it is plausible to assume that the frequent designation of SARS-CoV-2 variants as distinct entities, based on transient mutational changes rather than sustained functional divergence, represents a departure from well-established evolutionary virology principles.

Thus, the present study argues that the currently adopted classification of SARS-CoV-2 lineages has introduced significant challenges in COVID-19 risk communication. The existing SARS-CoV-2 classification systems are often the source of unnecessarily amplified public concern and inconsistent scientific justification that influences policy decisions, leading to a cycle of reactionary responses rather than well-balanced epidemiological measures. Accordingly, the present study aimed to: (1) Assess the genetic divergence of SARS-CoV-2 variants in comparison to established RNA virus speciation models (e.g., HIV-1,

HCV, and influenza A virus); (2) estimate the minimum evolutionary timeframe required for SARS-CoV-2 to reach a level of genetic divergence equivalent to that observed in HIV-1, HCV, and influenza A virus; and (3) propose a complementary classification framework for SARS-CoV-2 that is based on evolutionary and phylogenetic principles.

2. Materials and methods

2.1. Study design, dataset acquisition, and sequence selection

This study aimed to propose a practical and evolutionarily informed classification framework for SARS-CoV-2, grounded in a conservative, objective threshold of genetic divergence that could support the identification of biologically distinct lineages as the virus continues to evolve. To achieve this, the evolutionary divergence of SARS-CoV-2 variants was compared with that of three well-characterized RNA viruses – HIV-1, HCV, and influenza A virus. These viruses were selected as their lineage differentiation is associated with substantial biological and functional differences in transmissibility, immune evasion, and pathogenesis.

The viral genomes and corresponding genes chosen for analysis were: (1) SARS-CoV-2: *S* gene (3,822 bases); (2) HIV-1: *env* gene (2,592 bases); (3) HCV: *E1* gene (576 bases); and (4) influenza A virus: *HA* gene (1,707 bases). These regions were selected due to their relevance in viral entry, immune evasion, and vaccine targeting, besides being recognized as rapidly evolving genomic regions of each virus.[80-83] Full-length viral sequences were retrieved from the Los Alamos HIV Database, Los Alamos HCV Database, GenBank Influenza Virus Database, and NCBI Virus Database.[84-87] Selection criteria included: (1) Complete coding sequences (removal of sequences with large gaps or ambiguous bases, defined as >1% nucleotide base); (2) diversity in collection date and geographical origin to ensure representation of global viral evolution; (3) exclusion of recombinant sequences, except where recombination is inherent to viral evolution (e.g., HIV-1 circulating recombinant forms [CRFs]); and (4) at least 40 representative sequences per viral subtype/lineage to ensure robust phylogenetic reconstruction.[88] The selection of 50 (SARS-CoV-2), 58 (HIV-1), 40 (HCV), and 49 (influenza A virus) sequences adhered to this standard.

2.2. Selection of viral subtypes/lineages

To ensure a comprehensive comparison of SARS-CoV-2 genetic divergence with well-characterized RNA viruses, representative subtypes and lineages for HIV-1, HCV, and influenza A virus were selected based on their global prevalence and established phylogenetic classification. For SARS-CoV-2, sequences from the major VOCs were included in the analysis. These variants were selected based on their distinct S protein mutations and their role in altering transmissibility, immune escape, and vaccine efficacy.

HIV-1 exhibits significant genetic diversity, with distinct subtypes and CRFs contributing to its global dissemination. The following subtypes/CRFs, representing the most common lineages, were selected: A1, B, C, D, CRF01_AE, and CRF02_AG.[89] These subtypes were chosen due to their high epidemiological relevance and established divergence patterns, which serve as a benchmark for evaluating the evolutionary dynamics of SARS-CoV-2.[90]

HCV classification is based on well-defined genotypes and subtypes, which exhibit substantial genetic diversity.[91] The following subtypes were selected: subtype 1a, subtype 1b, subtype 3a, and subtype 4a. These subtypes represent widely studied lineages with distinct evolutionary trajectories, providing an appropriate comparison for assessing SARS-CoV-2 divergence. Influenza A virus evolves through antigenic drift and shift, leading to the emergence of distinct subtypes with significant genetic and antigenic variation. The following subtypes were included: H1N1 and H3N2.[92] These subtypes were selected as they represent the dominant circulating influenza lineages over the past century, providing a well-documented model of viral evolution and antigenic variation. For SARS-CoV-2, sequences from the five major VOCs designated by the WHO were analyzed, representing key evolutionary lineages that have dominated transmission waves globally: Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529).[43]

2.3. Sequence alignment and genetic distance estimation

Multiple sequence alignments were performed separately for HIV-1, HCV, influenza A, and SARS-CoV-2 using MAFFT v7.490, employing the L-INS-i algorithm, which optimizes alignment accuracy for sequences with frequent insertions and deletions, particularly in the HIV-1 *env* and influenza *HA* genes.[93] Alignment quality was manually reviewed in MEGA6, and sequences with poor alignment were excluded.[94] Genetic distances were estimated using the maximum composite likelihood (MCL) model, with rate variation among sites modeled using a gamma distribution (shape parameter = 1). Codon positions included 1st, 2nd, and 3rd coding positions, as well as non-coding regions.[94,95] All positions containing gaps or missing data were removed before analysis. The final alignment of HIV-1 (n = 58) consisted of 2,214 nucleotide positions; HCV (n = 40) consisted of 513 nucleotide positions; influenza

A virus (n = 49) consisted of 939 nucleotide positions; and SARS-CoV-2 (n = 50) consisted of 3,089 nucleotide positions. Evolutionary analyses were conducted using MEGA6 to compute pairwise genetic distances across all sequence groups.[94]

To enhance the temporal signal in assessing the genetic divergence of the viruses, we included sequences spanning multiple years and epidemic phases as follows. For HIV-1, the sequence collection years spanned 1993 – 2017, from 20 countries, including Argentina, Australia, Brazil, Botswana, China, Cyprus, Spain, Ethiopia, Finland, Indonesia, India, Iran, Kenya, Nigeria, Sweden, Thailand, Tanzania, Uganda, UK, and US. For HCV, the sequence collection years spanned 1993 – 2023, from 15 countries, including Australia, Switzerland, China, Cuba, Germany, Egypt, Spain, France, Ireland, India, Japan, Pakistan, Thailand, the UK, and the US. For influenza A, sequences were collected in multiple locations in Sweden spanning 1992 – 2010. For SARS-CoV-2, sequences were collected during 2020 – 2025 from Canada, Ghana, Japan, New Zealand, and several locations in the US, including Arizona, California, the District of Columbia, Iowa, Indiana, Michigan, Minnesota, North Carolina, Nevada, New Jersey, North Carolina, Oklahoma, Oregon, Pennsylvania, South Carolina, South Carolina, and Washington.

2.4. Maximum likelihood phylogenetic analysis

To assess evolutionary relationships and determine whether SARS-CoV-2 variants exhibit lineage divergence comparable to that observed in HIV-1, HCV, and influenza A virus, maximum likelihood (ML) phylogenetic trees were constructed using MEGA6.[94,95] The MCL model was employed for nucleotide substitution, with rate variation among sites modeled using a gamma distribution (shape parameter = 1). The analysis included all nucleotide sequences, with codon positions (first, second, third, and non-coding regions) considered. Sequences containing gaps or missing data were removed before analysis.[94,95] An unrooted ML tree was generated with 100 ultrafast bootstrap replicates to evaluate branch support, with bootstrap values exceeding 70% considered statistically significant.[96] The resulting tree was visualized using FigTree software to facilitate the interpretation of lineage relationships.[97] The inclusion of ≥40 sequences per virus provided sufficient phylogenetic resolution.[96,98]

2.5. Defining the genetic divergence threshold for variant classification

To establish an objective cutoff for defining distinct viral variants, the genetic distances of SARS-CoV-2 were compared to the following benchmarks: (1) HIV-1 subtypes (*env* region) mean divergence; (2) HCV subtypes (*E1* region) mean divergence; and (3) influenza A virus subtypes (*HA* gene) mean divergence. ANOVA and Kruskal–Wallis tests were used to compare mean genetic divergence values among viral groups, and the analysis was conducted using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp, United States). It was hypothesized that if SARS-CoV-2 variants do not surpass the established genetic divergence thresholds derived from HIV-1, HCV, and influenza A virus, this would indicate that their present classification is driven more by transient mutations than by meaningful virological differentiation.

2.6. Estimation of speciation time for SARS-CoV-2 based on genetic distances compared to HIV-1, HCV, and influenza A virus

To estimate the time required for SARS-CoV-2 to reach speciation-level divergence, we compared its genetic distances to those observed in HIV-1, HCV, and influenza A virus. Speciation thresholds were defined based on the minimum genetic distances observed between recognized subtypes or genotypes in these viruses. Using the established evolutionary rate of SARS-CoV-2 (0.0004 – 0.002 s/s/y), we applied the formula: Years = Genetic distance threshold/evolutionary rate. The evolutionary rate of SARS-CoV-2 was set at 0.0004 – 0.002 s/s/y, consistent with published estimates.[17,51,99-103] To evaluate how recombination affects divergence, an adjusted evolutionary rate model was incorporated based on the estimated recombination frequency in SARS-CoV-2 genomes (~2.7% recombinant ancestry).[104,105] A 1.5× acceleration factor was applied, as recombination has been shown to elevate the viral evolutionary rate,[106,107] yielding adjusted evolutionary rates: lower bound, 0.0006 s/s/y; upper bound, 0.003 s/s/y. Monte Carlo simulation was performed by generating 1,000 random evolutionary rates sampled uniformly from the adjusted range. Simulated rate values and calculated time estimates were compiled into a structured dataset and analyzed using the IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp, United States). Descriptive statistics (mean, standard deviation, and 95% confidence intervals [CIs]) were computed for the estimated time required to reach each threshold.

2.7. Basis of the proposal of new SARS-CoV-2 classification scheme

To establish a robust and biologically meaningful classification system for SARS-CoV-2 variants, a two-pronged methodological approach was employed, integrating (1) genetic divergence thresholds derived from well-characterized viral evolution patterns, and (2) functional impact criteria identified through a comprehensive review of literature on viral pathogenesis,

transmissibility, and immune escape mechanisms. Genetic divergence thresholds were established by analyzing the evolutionary distances between recognized subtypes or genotypes of HIV-1, HCV, and influenza A virus, as outlined in the sections above.

A review of the literature was conducted in PubMed/MEDLINE to identify key functional parameters that have historically defined viral lineages in other RNA viruses, including HIV-1, HCV, influenza A virus, and previous SARS-CoV-2 designated as VOCs. The exact search strategy, concluded on 31 January 2025, was: (("SARS-CoV-2" OR "COVID-19" OR "HIV-1" OR "Hepatitis C Virus" OR "HCV" OR "Influenza A") AND ("variant of concern" OR "VOC" OR "subtype" OR "genotype" OR "lineage" OR "clade") AND ("transmissibility" OR "replication fitness" OR "R0" OR "immune escape" OR "neutralizing antibodies" OR "vaccine resistance" OR "vaccine escape" OR "clinical severity" OR "hospitalization" OR "mortality" OR "epidemiological dominance" OR "variant persistence" OR "lineage replacement")).

Criteria were categorized into four virologically relevant domains: (1) Transmissibility: Evidence of increased replication fitness[40,91,108-112]; (2) Immune escape potential: ≥2-fold reduction in neutralizing antibody effectiveness, impacting vaccine efficacy, or convalescent immunity[92,113]; (3) Clinical impact: Increased hospitalization, mortality, or multisystem complications compared to prior circulating variants[114-119]; and (4) Epidemiological dominance: Variant must persist for >12 months, indicating sustained selective advantage.[89,120-122]

3. Results

3.1. HIV-1 *env* dataset description and ML phylogenetic analysis

A total of 58 full-length HIV-1 *env* sequences were analyzed, representing six major subtypes and CRFs. The ML phylogenetic tree robustly supported the classification of these subtypes into distinct lineages, with bootstrap values exceeding 0.99, indicating high confidence in the inferred evolutionary relationships (Figure 1). The distances ranged from 0.157 s/s (A vs. CRF02_AG) to 0.235 s/s (D vs. CRF01_AE), with an overall mean divergence of 0.21 s/s. The highest divergence was observed between subtypes D and CRF01_AE (0.235 s/s), whereas the lowest was found between subtypes A and CRF02_AG (0.157 s/s) (Table 1).

3.2. HCV *E1* dataset and evolutionary distance estimation

A total of 40 HCV *E1* sequences were analyzed, representing four major subtypes: 1a, 1b, 3a, and 4a. The dataset was curated to ensure full-length coding sequences, with all ambiguous bases and incomplete sequences removed before analysis. The final alignment contained a total of 2,214 nucleotide positions, with discrete clustering observed in the ML phylogenetic analysis (Figure 2).

The evolutionary divergence between subtypes was assessed using pairwise genetic distances (Table 2). The genetic distances among HCV subtypes ranged from 0.371 s/s (1a vs. 1b) to 0.674 s/s (3a vs. 4a), reflecting the well-documented high genetic diversity of HCV. The mean divergence across all subtype comparisons was 0.57 s/s, which is notably higher than that observed in HIV-1 *env* sequences.

3.3. Influenza A virus dataset and evolutionary distance analysis

The influenza dataset included 49 full-length *HA* sequences, representing two major subtypes, H1 and H3, as indicated in the ML phylogenetic tree (Figure 3). The genetic divergence between these subtypes was estimated at 1.956 s/s, which is substantially higher than the divergence observed in both HIV-1 and HCV subtypes.

3.4. SARS-CoV-2 dataset and evolutionary distance analysis

The SARS-CoV-2 *S* sequence dataset in the ML phylogenetic analysis showed a clear separation into five different statistically supported monophyletic clades (Figure 4). Pairwise genetic distance analysis of SARS-CoV-2 variants revealed minimal divergence, with nucleotide substitution rates ranging from 0.002 to 0.006 s/s. The highest genetic distance was observed between Omicron and Beta (0.006 s/s) and Omicron and Alpha (0.006 s/s), while the lowest was between Beta and Gamma (0.002 s/s). Variants Delta, Beta, and Gamma exhibited similar divergence levels relative to Alpha (0.003 – 0.004 s/s), indicating limited evolutionary separation. These findings demonstrate that SARS-CoV-2 variants remain within a constrained genetic space, with divergence levels substantially lower than those observed in HIV-1 subtypes, HCV genotypes, or influenza A virus subtypes (Table 3).

3.5. Establishing the genetic divergence threshold based on sequences analyzed

To objectively classify viral genetic divergence, three evolutionary divergence thresholds were established based on the minimum observed genetic distances in HIV-1, HCV, and influenza A virus. Three thresholds, shown in Figure 5, are as follows: low divergence, below 0.157 s/s (based on HIV-1 subtype minimum); moderate divergence, between 0.157 and 0.371 s/s (based on HCV subtype minimum); and high divergence, above 0.371 s/s (with influenza A virus defining the upper bound at

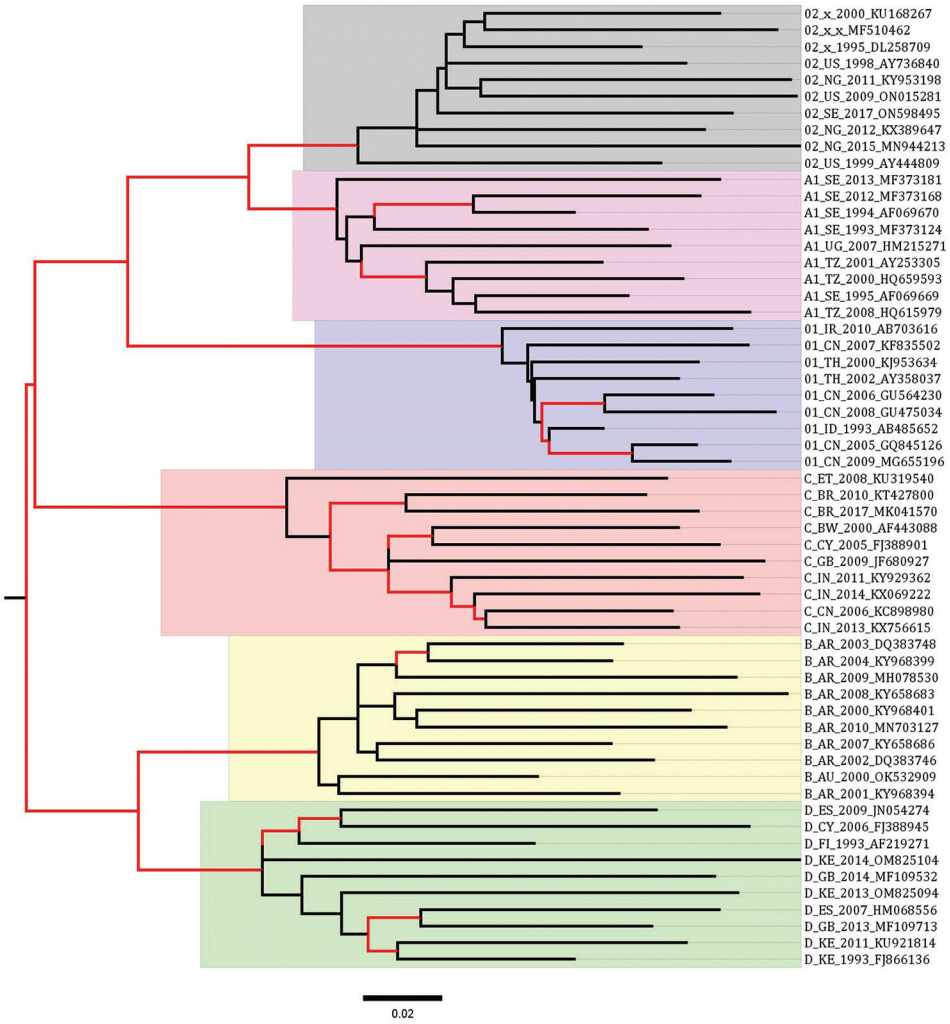

Figure 1. Maximum likelihood phylogenetic tree of HIV-1 subtypes/CRFs. Statistically supported branches are highlighted in red. The tree was visualized using the FigTree software.
Abbreviations: CRF: Circulating recombinant form; HIV-1: Human immunodeficiency virus 1.

1.956 s/s). HCV exhibited the highest mean genetic distance (0.5747 s/s, 95% CI: 0.4619 – 0.6874), consistent with the established differentiation of HCV genotypes. HIV-1 displayed moderate divergence (0.2159 s/s, 95% CI: 0.2043 – 0.2274), reflecting known subtype diversification. SARS-CoV-2 demonstrated extremely low divergence (0.004 s/s, 95% CI: 0.00299 – 0.00501), with all pairwise genetic distances ranging from 0.002 to 0.006 s/s, placing it well within the low divergence threshold. One-way ANOVA revealed highly significant differences in mean genetic distances among viral groups ($F = 606.821$, $p<0.001$), confirming that SARS-CoV-2 exhibits markedly lower divergence compared to HIV-1 and HCV. Kruskal–Wallis test further confirmed significant divergence differences ($H = 26.75$, $df = 3$, $p<0.001$).

3.6. Speciation time for SARS-CoV-2 based on genetic distances compared to HIV-1, HCV, and influenza A virus

Using 1,000 Monte Carlo-simulated evolutionary rates sampled uniformly from the adjusted range of 0.0006 to 0.003 s/s/y – to reflect SARS-CoV-2's empirically observed substitution rate, modified to account for recombination – we estimated the time required for the virus to reach genetic divergence thresholds comparable to those defining well-established subtypes of HIV-1, HCV, and influenza A virus.

For the HIV-1 divergence benchmark of 0.157 s/s, the estimated mean time for SARS-CoV-2 to reach an equivalent level of genetic divergence was 53.7 years (95% CI: 52.0 – 55.3), corresponding to a projected speciation

year of 2074 (range: 2046 – 2148). The median estimate was 43.8 years, or the year 2064, with a right-skewed distribution reflecting variability in substitution rates. For HCV-level divergence (0.371 s/s), the mean estimated timeframe was 126.8 years (95% CI: 123.0 – 130.6), corresponding to a projected divergence year of 2147 (range: 2082 – 2324), with a median of 103.5 years (year 2124). For the influenza A virus threshold (1.956 s/s), the mean estimated time to comparable divergence was 668.6 years (95% CI: 648.4 – 688.7), translating to a projected year of 2689 (range: 2346 – 3620), with a median of 545.7 years, or the year 2566 (Figure 6).

Table 1. Estimates of evolutionary divergence over sequence pairs between HIV-1 subtypes/CRFs

HIV-1 species 1	HIV-1 species 2	Genetic distance (s/s)
CRF01_AE	C	0.231
CRF02_AG	CRF01_AE	0.200
CRF02_AG	C	0.227
A	CRF02_AG	0.157
A	CRF01_AE	0.203
A	C	0.223
B	D	0.188
B	A	0.220
B	CRF02_AG	0.223
B	CRF01_AE	0.226
B	C	0.227
D	A	0.219
D	CRF02_AG	0.230
D	CRF01_AE	0.235
D	C	0.229

Abbreviations: CRFs: Circulating recombinant forms; HIV-1: Human immunodeficiency virus 1; s/s: Substitutions per site.

3.7. Application of the new literature-based classification system to existing SARS-CoV-2 variants

To evaluate whether existing SARS-CoV-2 variants met the criteria for variant classification, each was assessed using the new literature-based classification system as follows: (1) Genetic distance analysis: None of the examined SARS-CoV-2 variants exceeded the moderate divergence threshold (0.157 s/s); (2) Functional and epidemiological review: While some variants demonstrated increased transmissibility and immune escape, none met all classification functional criteria. Omicron sub-lineages exhibited notable immune evasion but did not substantially increase clinical severity or hospitalization rates beyond prior VOCs; and (3) Longitudinal surveillance: No SARS-CoV-2 variant maintained dominance for >12 months or necessitated a fundamental vaccine update, which would indicate a stable antigenic drift pattern. These findings suggest that the present classification of SARS-CoV-2 variants as distinct lineages is not supported by genetic, virological, or epidemiological data. Under

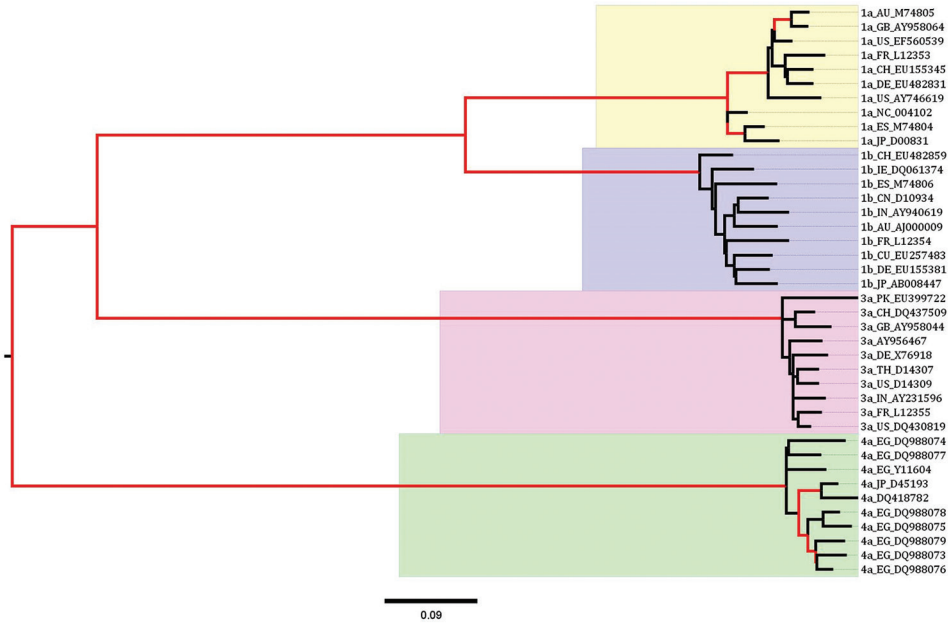

Figure 2. Maximum likelihood phylogenetic tree of HCV subtypes. Statistically supported branches are highlighted in red. The tree was visualized using the FigTree software.
Abbreviation: HCV: Hepatitis C virus.

the suggested classification system, none of the examined SARS-CoV-2 variants qualify as VOCs by both genetic and functional impact criteria, reinforcing the need for a revised classification framework based on objective thresholds.

Table 2. Estimates of evolutionary divergence over sequence pairs between HCV subtypes

HCV species 1	HCV species 2	Genetic distance (s/s)
1a	1b	0.371
1a	3a	0.613
1b	3a	0.554
1a	4a	0.599
1b	4a	0.637
3a	4a	0.674

Abbreviations: HCV: Hepatitis C virus; s/s: Substitutions per site.

4. Discussion

Although the present SARS-CoV-2 classification schemes played a role in early variant detection and response planning, several shortcomings remain that could be addressed with a complementary evolutionary and biologically based system. First, the present schemes overgeneralize risk, frequently triggering disproportionate public reactions before sufficient epidemiological data are available. Second, they lack robust quantitative severity metrics, often prioritizing genetic alterations over concrete clinical outcomes or real-world vaccine effectiveness. Third, they inadvertently amplify public anxiety by reinforcing the notion that each new variant constitutes an existential crisis, rather than a predictable feature of viral evolution.[123,124] A shift toward a refined, biologically relevant classification system is therefore needed. Rather than

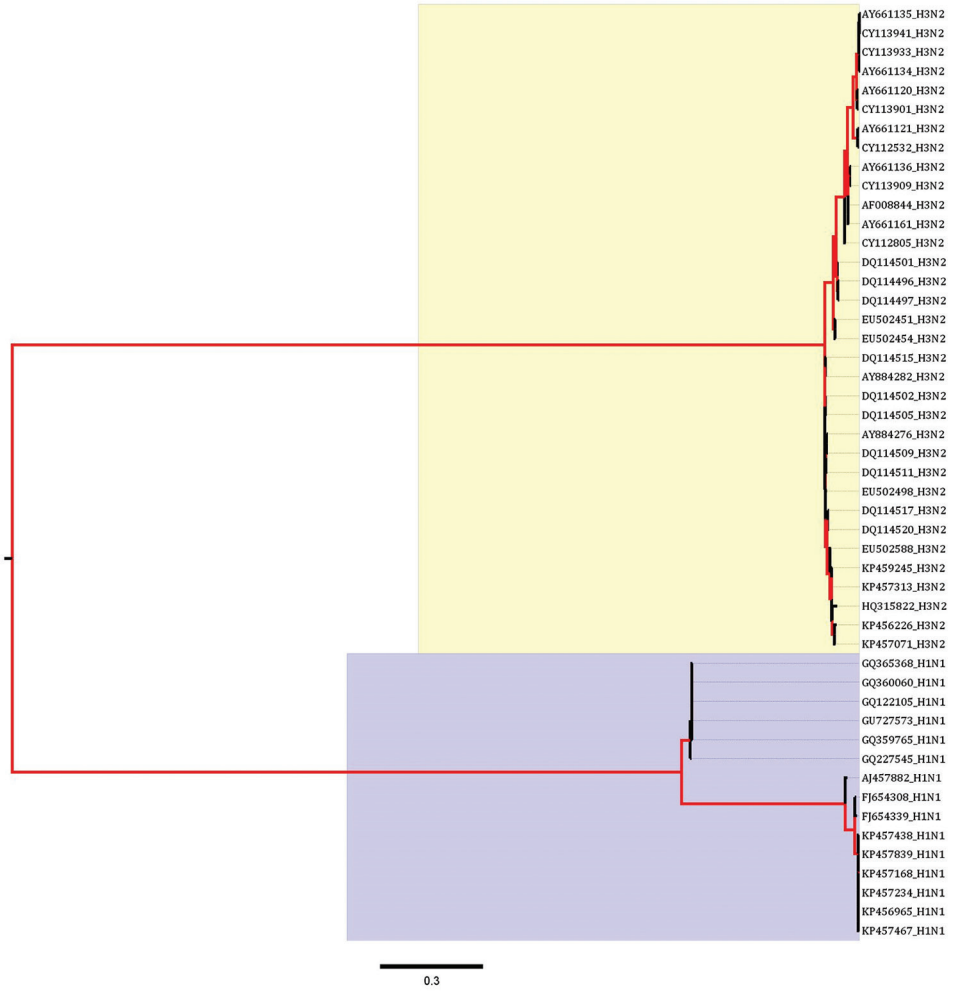

Figure 3. Maximum likelihood phylogenetic tree of influenza A virus subtypes. Statistically supported branches are highlighted in red. The tree was visualized using the FigTree software.

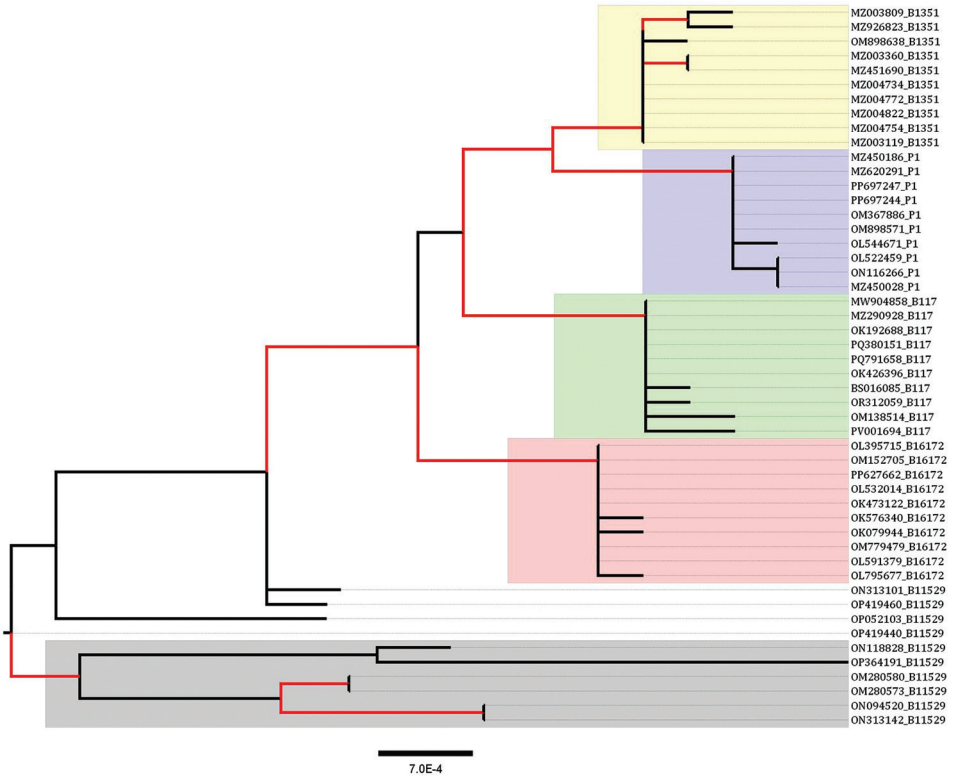

Figure 4. Maximum likelihood phylogenetic tree of SARS-CoV-2 VOCs. Statistically supported branches are highlighted in red. The tree was visualized using the FigTree software.
Abbreviations: SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; VOCs: Variants of concern.

Table 3. Estimates of evolutionary divergence over sequence pairs between SARS-CoV-2 VOCs

SARS-CoV-2 species 1	SARS-CoV-2 species 2	Genetic distance (s/s)
Omicron	Delta	0.005
Omicron	Beta	0.006
Delta	Beta	0.003
Omicron	Gamma	0.005
Delta	Gamma	0.003
Beta	Gamma	0.002
Omicron	Alpha	0.006
Delta	Alpha	0.003
Beta	Alpha	0.003
Gamma	Alpha	0.004

Abbreviations: SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; s/s: Substitutions per site; VOCs: Variants of concern.

abandoning existing surveillance efforts, a new approach would re-anchor SARS-CoV-2 variant designation in a long-term phylogenetic context – improving public health communication, reducing overreaction, and aligning public health policy with actual risk.

The recurrent emergence of new SARS-CoV-2 variants has repeatedly led to heightened concern. These concerns were amplified by media coverage with reactive health policy shifts despite limited evidence of substantial biological divergence of these variants.[125] From Alpha to Omicron and its subsequent sub-lineages, the actual epidemiological impact of these variants has exhibited substantial variability.[126,127] While certain mutations have been associated with enhanced transmissibility, immune evasion, or altered pathogenicity, the magnitude of these changes is often overstated in public discussions.[114,128-130] The assumption that every newly designated variant introduces a profound shift in SARS-CoV-2 behavior lacks consistent empirical validation. Most SARS-CoV-2 variants remain genetically and phenotypically constrained within a narrow evolutionary space, exhibiting only minor incremental changes rather than the transformative antigenic shifts observed in viruses such as influenza A virus.[131,132]

The classification of SARS-CoV-2 variants has dominated scientific discourse and public health decision-making, yet findings from the present study reveal fundamental flaws in the system currently in use. The

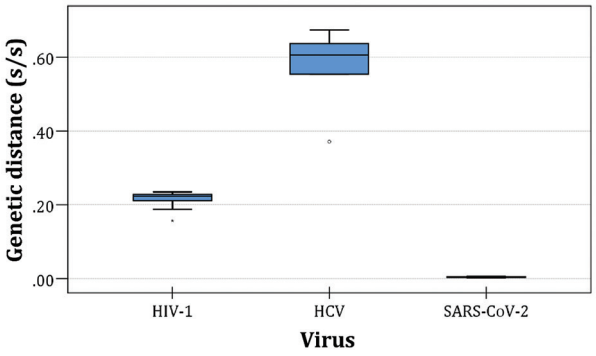

Figure 5. Genetic distance distributions among HIV-1, HCV, and SARS-CoV-2
Notes: Circles indicate mild outliers (1.5 – 3× interquartile range (IQR) from Q1 or Q3); asterisks indicate extreme outliers (>3× IQR).
Abbreviations: HCV: Hepatitis C virus; HIV-1: Human immunodeficiency virus 1; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; s/s: Substitutions per site.

existing system, which rapidly designates emerging lineages as VOCs based on relatively minor genetic mutations, lacks quantifiable thresholds rooted in evolutionary virology. This study establishes a comparative framework by analyzing genetic divergence in well-characterized RNA viruses, including HIV-1, HCV, and influenza A virus, to determine whether SARS-CoV-2 variants exhibit levels of genetic differentiation that justify their classification as distinct viral entities. Results indicated that SARS-CoV-2 genetic distances are orders of magnitude below those observed in these benchmark viruses, demonstrating that the repeated designation of SARS-CoV-2 variants has been driven more by reactionary classification than by true viral evolution.

The classification of RNA viruses into lineages, variants, subtypes, clades, or genotypes has historically been based on genetic divergence thresholds that correlate with functional, epidemiological, or antigenic differentiation.[133-135] In this study, HIV-1 subtype classification was defined by a minimum genetic distance of 0.157 s/s, with clear distinctions in antigenicity, immune evasion, and drug resistance between HIV-1 subtypes/CRFs, as indicated by previous studies.[136-138] In addition, HCV subtypes exhibited even greater divergence (≥0.371 s/s), with genetic differences that impact both viral pathogenesis and treatment responses, as indicated by a multitude of previous studies.[139-142] Influenza A virus subtypes, which undergo antigenic shift, demonstrate a genetic divergence of approximately 1.956 s/s, corresponding to profound structural and functional changes in the hemagglutinin (HA) protein that drive immune escape and reinfection dynamics.[143,144]

In contrast, the analysis revealed that SARS-CoV-2 variants exhibited a maximum genetic distance of only

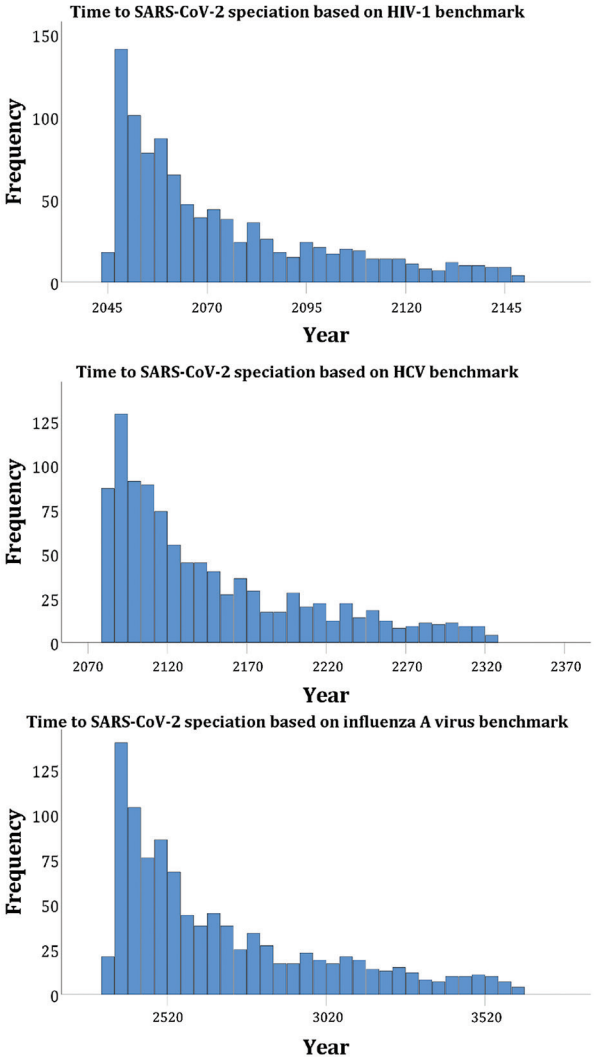

Figure 6. Projected time required for SARS-CoV-2 to reach speciation-level divergence based on comparative viral benchmarks
Abbreviations: HCV: Hepatitis C virus; HIV-1: Human immunodeficiency virus 1; SARS-CoV-2: Severe acute respiratory syndrome coronavirus.

0.006 s/s, placing them significantly below the minimum divergence threshold established for the HIV-1 subtype. This indicates that the widely publicized SARS-CoV-2 variants, including Alpha, Delta, and Omicron, fall within the range of normal intra-lineage variation rather than approaching the level of divergence that would justify classification as separate viral lineages. The absence of substantial genetic separation challenges the rationale behind the present SARS-CoV-2 classification system. Emerging data from the later stages of the COVID-19 pandemic (2023 – 2024) support the view that SARS-CoV-2 continues to evolve within a constrained genetic landscape.[145] For example, Markov *et al.*[49] described the appearance of VOCs as "shift-

like" events – abrupt, mutation-rich transitions following prolonged periods of undetected evolution. These contrast with the more gradual, "drift-like" evolution observed within the Omicron lineage.[49] Complementary CDC genomic surveillance during 2023 – 2024 indicated that dominant Omicron sub-lineages (e.g., XBB.1.5, JN.1) remained genetically close to earlier variants.[146,147]

To further contextualize the slow evolutionary trajectory of SARS-CoV-2, the time required for the virus to reach genetic divergence thresholds observed in HIV-1, HCV, and influenza A was modeled. Using the estimated evolutionary rate of SARS-CoV-2 (0.0004 – 0.002 s/s per year), the timeframes necessary for the accumulation of genetic distances comparable to those of other RNA viruses were calculated as follows: On average, HIV-1-like divergence would require 54 years; HCV-like divergence would require 127 years; and influenza A virus-like divergence would require 669 years. Even under accelerated evolutionary scenarios that incorporate recombination events, the time required for SARS-CoV-2 to evolve into a distinct viral lineage is projected to exceed the present pandemic timeline. This finding disputes the variant-driven narrative that suggests SARS-CoV-2 is undergoing rapid evolutionary shifts with each new lineage designation. Instead, the data support a preliminary conclusion that SARS-CoV-2 remains evolutionarily constrained, undergoing relatively slow genetic drift rather than true speciation.

As stated earlier, the WHO classification framework for SARS-CoV-2 variants is not entirely based on established genetic divergence thresholds but rather on epidemiological trends and qualitative assessments of transmissibility, immune escape, and severity.[35] However, the findings of this study suggest several contradictory aspects. Variant designations are applied inconsistently, with transient genetic mutations (e.g., D614G, N501Y) serving as the primary justification for labeling variants, despite similar mutations occurring in other viruses without warranting separate classification.[45,148]

Further supporting the hypothesis that some SARS-CoV-2 variant classifications may lack robust biological justification is the consistent observation that functional changes attributed to SARS-CoV-2 variants are quantitatively minor, particularly in comparison to antigenic shifts in influenza A virus.[75,149] While Omicron sub-lineages exhibit a 3 – 4-fold reduction in neutralizing antibody titers, this pales in comparison to the differences observed between HIV-1 subtypes/CRFs, highlighting the exaggerated portrayal of SARS-CoV-2 mutations as profound evolutionary shifts.[150,151] Thus, the epidemiological dominance of a SARS-CoV-2 variant at a given time point was often misinterpreted as increased viral fitness, without sufficient consideration of immune landscape effects, stochastic founder effects, and shifting population-level immunity in shaping variant emergence.[43] By relying on short-term epidemiological observations rather than virological principles, the present classification system of SARS-CoV-2 VOCs appears to promote an illusion of continuous viral emergence, despite genetic evidence indicating relative evolutionary stability of SARS-CoV-2, as shown by Lv *et al.*[71]

A more rigorous and biologically grounded framework for classifying SARS-CoV-2 variants would complement the existing system by providing clearer genetic and functional thresholds. This integrative approach could enhance the present SARS-CoV-2 classification schemes, offering additional context for long-term surveillance while preserving their relevance for rapid public health response. This approach would help to ensure that SARS-CoV-2 variant classifications are based on measurable evolutionary changes, rather than transient mutations that have contributed to public anxiety and misdirected public health responses.[123,124] A robust classification system must first define genetic divergence thresholds aligned with the established evolutionary dynamics of RNA viruses. In the case of HIV-1, HCV, and influenza A virus, genetic distances serve as a key determinant of whether a viral strain is classified as a distinct subtype/CRF, genotype/sub-genotype, or lineage.

While genetic divergence serves as the foundation of viral classification, the functional consequences of mutations must also be considered. A variant should not be designated solely based on genetic alterations, but rather on whether these changes confer a meaningful impact on transmission, immune evasion, or disease severity. Functional criteria must also be met in conjunction with genetic divergence thresholds to ensure that only biologically significant variants are designated as VOCs. These functional criteria include: (1) A demonstrable increase in transmissibility compared to the previously dominant variant; (2) a significant reduction in neutralizing antibody titers, confirmed in real-world immunological studies rather than isolated *in vitro* assays, indicating a meaningful immune escape advantage; (3) a significant increase in hospitalization rates attributable to increased intrinsic virulence, rather than confounding factors such as healthcare system burden or case misclassification; and (4) sustained epidemiological dominance lasting ≥12 months, ensuring that the emergent variant represents a sustained shift in viral evolution rather than a transient lineage fluctuation.

To put these criteria into practical aspects, influenza emerges as a stark example with antigenic shifts between

H1N1 and H3N2 being associated with near-complete immune escape and requiring substantial vaccine reformulation with a long period of circulation.[152,153] In contrast, no SARS-CoV-2 variant has maintained epidemiological dominance for over 12 months, a key criterion for distinguishing transient genetic shifts from long-term evolutionary changes. For instance, while Delta emerged as the dominant strain for several months, it was rapidly displaced by Omicron, which, in turn, has continued to generate multiple sub-lineages without a single, stable successor.[154] This pattern strongly suggests that SARS-CoV-2 evolution is characterized by alternating variant prevalence rather than sustained antigenic evolution, a hallmark of viruses undergoing immune-driven lineage differentiation. On the other hand, HIV-1 subtypes/CRFs, HCV genotypes/subtypes, and influenza A subtypes maintain long-term epidemiological dominance, persisting for decades despite selective pressures from immune responses and antiviral therapies.[138,141]

An important aspect that motivated this study can be presented as follows. By failing to incorporate both genetic divergence thresholds and functional impact assessments, the present WHO SARS-CoV-2 VOCs classification system may have resulted in unnecessary alarm and misdirected public health interventions.[155,156] The repeated reclassification of variants might have eroded public trust by creating the impression that SARS-CoV-2 is undergoing rapid, unpredictable evolutionary leaps, despite evidence indicating that it remains a genetically stable virus with a slow rate of change. This has also led to inefficient resource allocation, with disproportionate emphasis placed on variant-specific vaccine formulations, booster campaigns, and travel restrictions, rather than investing in broad-spectrum immunological strategies that provide durable protection against viral evolution.[47]

The potential misclassification of SARS-CoV-2 variants could have far-reaching consequences, influencing public health policies, economic stability, and scientific priorities, often in ways that lack a solid virological foundation. Reactive policy decisions, including travel bans, lockdowns, and stringent restrictions, were hastily implemented in response to the emergence of Beta, Gamma, and Omicron, despite these variants exhibiting genetic distances well within the bounds of normal intra-lineage variation.[157,158] Such responses were not only scientifically unjustified but also economically disruptive, fueling cycles of uncertainty that undermined long-term pandemic planning. Compounding this issue, the frequent and often sensationalized designation of new variants has significantly eroded public trust in health authorities. The portrayal of SARS-CoV-2 as an unpredictable, rapidly evolving virus – despite genomic evidence indicating relatively slow evolutionary change – has led to widespread misinformation, fostering public skepticism toward vaccination efforts and pandemic control measures.[159] In addition, scientific resources have been misallocated, with an excessive focus on sequencing minor genetic mutations rather than investing in broadly protective vaccines, long-term antiviral solutions, and cross-reactive immunological strategies. This relentless pursuit of tracking and naming every emerging sub-lineage has diverted focus from more sustainable public health interventions, including endemic virus management and the development of durable, pan-coronavirus strategies better suited for long-term mitigation efforts.

Building on the findings of this study, several areas for future research are warranted. Comparative genomic analyses should be expanded to include other coronaviruses – such as SARS-CoV-1, Middle East respiratory syndrome coronavirus (MERS-CoV), and endemic human coronaviruses – to establish divergence thresholds that are biologically tailored to the coronavirus family, rather than relying solely on benchmarks derived from HIV-1, HCV, and influenza A virus. Evolutionary rate modeling should be extended to longer timeframes, incorporating host adaptation, immune-driven selection, and antiviral pressure, all of which may alter divergence trajectories over time. Given the unique RNA proofreading mechanisms in coronaviruses, future work should also assess the role of recombination in shaping long-term evolution. Studies should investigate whether observed genetic divergence corresponds to antigenic drift by utilizing neutralization assays across multiple viral proteins beyond the spike. Moreover, large-scale meta-analyses are needed to clarify whether specific mutations consistently associate with increased clinical severity. Variant-specific models should be integrated into real-world vaccine effectiveness studies to better quantify the impact of mutations on breakthrough infections and the duration of immunity. Finally, the framework proposed in this study could be integrated with existing nomenclature systems such as Pango and the WHO variant classifications. By delineating transient, antigenically responsive sub-lineages from those exhibiting sustained phylogenetic divergence, this approach could provide a complementary strategy that improves evolutionary resolution and may help prevent overinterpretation of SARS-CoV-2 genomic changes, thereby reducing unnecessary public health interventions.

Finally, this study is subject to several limitations that should be taken into account as follows. First, SARS-CoV-2 has circulated in humans for a relatively short period, and while evolutionary projections were made based on the

present mutation rates, unforeseen selective pressures may alter these trajectories. Second, although SARS-CoV-2 has exhibited limited divergence to date, the potential for future recombination or adaptive mutations remains. Third, while the analysis included the most epidemiologically significant SARS-CoV-2 variants, it may not capture the evolutionary dynamics of rare or geographically restricted lineages. Fourth, evolutionary rate estimates are constrained by the availability and geographic distribution of sequencing data, which may overrepresent certain SARS-CoV-2 variants due to sequencing concentration in high-income countries. Fifth, although recombination was incorporated into the modeling, its true long-term impact on divergence is not fully understood. Sixth, inferences about transmissibility were drawn from epidemiological data rather than direct experimental comparisons, which were beyond the scope of this study. Finally, while the proposed classification system aims to guide future policy and research, its global adoption will require coordination and consensus among international public health authorities with potentially divergent priorities.

5. Conclusion

The classification of SARS-CoV-2 variants has profoundly shaped public health policy, vaccine development, and global perceptions throughout the pandemic. However, the analysis suggests that many of these classifications have not been consistently anchored in robust virological criteria, highlighting an opportunity to enhance the present system with more biologically grounded standards. Comparative genomic analyses demonstrate that the maximum divergence among SARS-CoV-2 variants (0.006 s/s) is substantially below the thresholds used to define subtypes in HIV-1, HCV, or influenza A virus. Even under accelerated evolutionary scenarios, the virus would require centuries to reach divergence levels warranting lineage designation. These findings challenge the prevailing narrative of rapid, unpredictable viral evolution and instead suggest that SARS-CoV-2 remains evolutionarily constrained. The frequent and inconsistent designation of VOCs has led to public confusion, reactionary policy decisions, and the diversion of scientific resources. To address these issues, an evidence-based framework integrating genetic, functional, and epidemiological criteria is proposed to classify SARS-CoV-2 variants. Under this proposed framework, none of the existing SARS-CoV-2 VOCs meet the combined genetic and functional thresholds typically required for lineage-level classification in other RNA viruses. Rather than replacing present classification systems, the proposed approach advocates for a paradigm shift that distinguishes persistent evolutionary patterns from short-term genetic fluctuations – to support more proportionate, evidence-based responses in future phases of the pandemic.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Malik Sallam
Data curation: All authors
Formal analysis: Saja Al-Baidhani, Malik Sallam
Investigation: All authors
Methodology: All authors
Project administration: Malik Sallam
Supervision: Malik Sallam
Validation: Saja Al-Baidhani, Malik Sallam
Visualization: Malik Sallam
Writing – original draft: Malik Sallam
Writing – review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data presented in this study are available on request from the corresponding author (Malik Sallam).

References

1. Gorbalenya AE, Baker SC, Baric RS, *et al*. The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5(4):536-544.

   doi: 10.1038/s41564-020-0695-z

2. Lu R, Zhao X, Li J, *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-574.

   doi: 10.1016/s0140-6736(20)30251-8

3. Zhou P, Yang XL, Wang XG, *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273.

   doi: 10.1038/s41586-020-2012-7

4. Rosen CJ. Viral variants, vaccinations, and long covid - new insights. *N Engl J Med*. 2024;391(6):561-562.

doi: 10.1056/NEJMe2407575

5. He F, Deng Y, Li W. Coronavirus disease 2019: What we know? *J Med Virol*. 2020;92(7):719-725.

   doi: 10.1002/jmv.25766

6. Joseph S, Kutty Narayanan A. Covid 19-the 21st century pandemic: The novel coronavirus outbreak and the treatment strategies. *Adv Pharm Bull*. 2022;12(1):34-44.

   doi: 10.34172/apb.2022.005

7. Sah P, Fitzpatrick MC, Zimmer CF, *et al*. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proc Natl Acad Sci U S A*. 2021;118(34):e2109229118.

   doi: 10.1073/pnas.2109229118

8. Gandhi RT, Lynch JB, Del Rio C. Mild or moderate Covid-19. *N Engl J Med*. 2020;383(18):1757-1766.

   doi: 10.1056/NEJMcp2009249

9. Çelik I, Öztürk R. From asymptomatic to critical illness: Decoding various clinical stages of COVID-19. *Turk J Med Sci*. 2021;51(Si-1):3284-3300.

   doi: 10.3906/sag-2107-137

10. Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. SARS-CoV-2 and COVID-19 in older adults: What we may expect regarding pathogenesis, immune responses, and outcomes. *GeroScience*. 2020;42(2):505-514.

    doi: 10.1007/s11357-020-00186-0

11. Boyton RJ, Altmann DM. The immunology of asymptomatic SARS-CoV-2 infection: What are the key questions? *Nat Rev Immunol*. 2021;21(12):762-768.

    doi: 10.1038/s41577-021-00631-x

12. Naqvi AAT, Fatima K, Mohammad T, *et al*. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim Biophys Acta Mol Basis Dis*. 2020;1866(10):165878.

    doi: 10.1016/j.bbadis.2020.165878

13. Rahimi A, Mirzazadeh A, Tavakolpour S. Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics*. 2021;113(1 Pt 2):1221-1232.

    doi: 10.1016/j.ygeno.2020.09.059

14. Alwine JC, Casadevall A, Enquist LW, Goodrum FD, Imperiale MJ. A critical analysis of the evidence for the SARS-CoV-2 origin hypotheses. *mSphere*. 2023;8(2):e0011923.

    doi: 10.1128/msphere.00119-23

15. Hao YJ, Wang YL, Wang MY, *et al*. The origins of COVID-19 pandemic: A brief overview. *Transbound Emerg Dis*. 2022;69(6):3181-3197.

    doi: 10.1111/tbed.14732

16. Boni MF, Lemey P, Jiang X, *et al*. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol*. 2020;5(11):1408-1417.

    doi: 10.1038/s41564-020-0771-4

17. Pereson MJ, Mojsiejczuk L, Martínez AP, Flichman DM, Garcia GH, Di Lello FA. Phylogenetic analysis of SARS-CoV-2 in the first few months since its emergence. *J Med Virol*. 2021;93(3):1722-1731.

    doi: 10.1002/jmv.26545

18. Canuti M, Bianchi S, Kolbl O, *et al*. Waiting for the truth: is reluctance in accepting an early origin hypothesis for SARS-CoV-2 delaying our understanding of viral emergence? *BMJ Glob Health*. 2022;7(3):e008386.

    doi: 10.1136/bmjgh-2021-008386

19. Pekar JE, Magee A, Parker E, *et al*. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science*. 2022;377(6609):960-966.

    doi: 10.1126/science.abp8337

20. Bukin YS, Bondaryuk AN, Kulakova NV, Balakhonov SV, Dzhioev YP, Zlobin VI. Phylogenetic reconstruction of the initial stages of the spread of the SARS-CoV-2 virus in the Eurasian and American continents by analyzing genomic data. *Virus Res*. 2021;305:198551.

    doi: 10.1016/j.virusres.2021.198551

21. Domingo JL. What we know and what we need to know about the origin of SARS-CoV-2. *Environ Res*. 2021;200:111785.

    doi: 10.1016/j.envres.2021.111785

22. Latinne A, Hu B, Olival KJ, *et al*. Origin and cross-species transmission of bat coronaviruses in China. *Nat Commun*. 2024;15(1):10705.

    doi: 10.1038/s41467-024-55384-7

23. Voskarides K. SARS-CoV-2: Tracing the origin, tracking the evolution. *BMC Med Genomics*. 2022;15(1):62.

    doi: 10.1186/s12920-022-01208-w

24. Lopes LR, de Mattos Cardillo G, Paiva PB. Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host. *Braz J Microbiol*. 2020;51(4):1593-1599.

    doi: 10.1007/s42770-020-00321-1

25. Samson S, Lord É, Makarenkov V. Assessing the emergence time of SARS-CoV-2 zoonotic spillover. *PLoS One*. 2024;19(4):e0301195.

    doi: 10.1371/journal.pone.0301195

26. Ruiz-Aravena M, McKee C, Gamble A, *et al*. Ecology, evolution and spillover of coronaviruses from bats. *Nat Rev Microbiol*. 2022;20(5):299-314.

    doi: 10.1038/s41579-021-00652-2

27. Holmes EC, Goldstein SA, Rasmussen AL, *et al*. The origins of

SARS-CoV-2: A critical review. *Cell*. 2021;184(19):4848-4856.

doi: 10.1016/j.cell.2021.08.017

28. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr Biol*. 2020;30(7):1346-1351.e2.

doi: 10.1016/j.cub.2020.03.022

29. Tang X, Ying R, Yao X, *et al*. Evolutionary analysis and lineage designation of SARS-CoV-2 genomes. *Sci Bull*. 2021;66(22):2297-2311.

doi: 10.1016/j.scib.2021.02.012

30. Rambaut A, Holmes EC, O'Toole Á, *et al*. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;5(11):1403-1407.

doi: 10.1038/s41564-020-0770-5

31. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci U S A*. 2020;117(17):9241-9243.

doi: 10.1073/pnas.2004999117

32. Lemey P, Hong SL, Hill V, *et al*. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat Commun*. 2020;11(1):5110.

doi: 10.1038/s41467-020-18877-9

33. Worobey M, Pekar J, Larsen BB, *et al*. The emergence of SARS-CoV-2 in Europe and North America. *Science*. 2020;370(6516):564-570.

doi: 10.1126/science.abc8169

34. Hill V, Du Plessis L, Peacock TP, *et al*. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol*. 2022;8(2):veac080.

doi: 10.1093/ve/veac080

35. World Health Organization. *Tracking SARS-CoV-2 Variants*; 2025. Available from: https://www.who.int/activities/tracking-SARS-CoV-2-variants [Last accessed on 2025 May 09].

36. Essabbar A, El Mazouri S, Boumajdi N, *et al*. Temporal dynamics and genomic landscape of SARS-CoV-2 after four years of evolution. *Cureus*. 2024;16(2):e53654.

doi: 10.7759/cureus.53654

37. Brüssow H. COVID-19: Emergence and mutational diversification of SARS-CoV-2. *Microb Biotechnol*. 2021;14(3):756-768.

doi: 10.1111/1751-7915.13800

38. Korber B, Fischer WM, Gnanakaran S, *et al*. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. 2020;182(4):812-827.e19.

doi: 10.1016/j.cell.2020.06.043

39. Sallam M, Ababneh NA, Dababseh D, Bakri FG, Mahafzah A. Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa. *Heliyon*. 2021;7(1):e06035.

doi: 10.1016/j.heliyon.2021.e06035

40. Earnest R, Uddin R, Matluk N, *et al*. Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. *Cell Rep Med*. 2022;3(4):100583.

doi: 10.1016/j.xcrm.2022.100583

41. Sallam M, Mahafzah A. Molecular analysis of SARS-CoV-2 genetic lineages in Jordan: Tracking the introduction and spread of COVID-19 UK variant of concern at a country level. *Pathogens*. 2021;10(3):302.

doi: 10.3390/pathogens10030302

42. Chouikha A, Fares W, Laamari A, *et al*. Molecular epidemiology of SARS-CoV-2 in Tunisia (North Africa) through several successive waves of COVID-19. *Viruses*. 2022;14(3):624.

doi: 10.3390/v14030624

43. Carabelli AM, Peacock TP, Thorne LG, *et al*. SARS-CoV-2 variant biology: Immune escape, transmission and fitness. *Nat Rev Microbiol*. 2023;21(3):162-177.

doi: 10.1038/s41579-022-00841-7

44. Pondé RAA. Physicochemical effect of the N501Y, E484K/Q, K417N/T, L452R and T478K mutations on the SARS-CoV-2 spike protein RBD and its influence on agent fitness and on attributes developed by emerging variants of concern. *Virology*. 2022;572:44-54.

doi: 10.1016/j.virol.2022.05.003

45. Perez-Gomez R. The development of SARS-CoV-2 variants: The gene makes the disease. *J Dev Biol*. 2021;9(4):58.

doi: 10.3390/jdb9040058

46. Harris E. WHO launches broad coronavirus surveillance network. *JAMA*. 2024;331(19):1615.

doi: 10.1001/jama.2024.6724

47. Otto SP, Day T, Arino J, *et al*. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Curr Biol*. 2021;31(14):R918-R929.

doi: 10.1016/j.cub.2021.06.049

48. O'Toole Á, Pybus OG, Abram ME, Kelly EJ, Rambaut A. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics*. 2022;23(1):121.

doi: 10.1186/s12864-022-08358-2

49. Markov PV, Ghafari M, Beer M, *et al*. The evolution of SARS-CoV-2. *Nat Rev Microbiol*. 2023;21(6):361-379.

doi: 10.1038/s41579-023-00878-2

50. Shahhosseini N, Babuadze GG, Wong G, Kobinger GP. Mutation signatures and *in silico* docking of novel SARS-CoV-2

variants of concern. *Microorganisms*. 2021;9(5):926.

doi: 10.3390/microorganisms9050926

51. Tao K, Tzou PL, Nouhin J, *et al*. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet*. 2021;22(12):757-773.

doi: 10.1038/s41576-021-00408-x

52. Subissi L, Otieno JR, Worp N, *et al.* An updated framework for SARS-CoV-2 variants reflects the unpredictability of viral evolution. *Nat Med*. 2024;30(9):2400-2403.

doi: 10.1038/s41591-024-02949-0

53. Giovanetti M, Benedetti F, Campisi G, *et al*. Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochem Biophys Res Commun*. 2021;538:88-91.

doi: 10.1016/j.bbrc.2020.10.102

54. Hussain B, Wu C. Evolutionary and phylogenetic dynamics of SARS-CoV-2 variants: A genetic comparative study of Taiyuan and Wuhan Cities of China. *Viruses*. 2024;16(6):907.

doi: 10.3390/v16060907

55. Zabidi NZ, Liew HL, Farouk IA, *et al.* Evolution of SARS-CoV-2 variants: Implications on immune escape, vaccination, therapeutic and diagnostic strategies. *Viruses*. 2023;15(4):944.

doi: 10.3390/v15040944

56. Kumar R, Srivastava Y, Muthuramalingam P, *et al*. Understanding mutations in human SARS-CoV-2 spike glycoprotein: A systematic review & meta-analysis. *Viruses*. 2023;15(4):856.

doi: 10.3390/v15040856

57. Bhattacharya M, Chatterjee S, Lee SS, Dhama K, Chakraborty C. Antibody evasion associated with the RBD significant mutations in several emerging SARS-CoV-2 variants and its subvariants. *Drug Resist Updat*. 2023;71:101008.

doi: 10.1016/j.drup.2023.101008

58. Liu Y, Liu J, Plante KS, *et al*. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*. 2022;602(7896):294-299.

doi: 10.1038/s41586-021-04245-0

59. Oude Munnink BB, Worp N, Nieuwenhuijse DF, *et al*. The next phase of SARS-CoV-2 surveillance: Real-time molecular epidemiology. *Nat Med*. 2021;27(9):1518-1524.

doi: 10.1038/s41591-021-01472-w

60. Parums V. Editorial: Revised world health organization (WHO) terminology for variants of concern and variants of interest of SARS-CoV-2. *Med Sci Monit*. 2021;27:e933622.

doi: 10.12659/msm.933622

61. Chen Z, Azman AS, Chen X, *et al*. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nat Genet*. 2022;54(4):499-507.

doi: 10.1038/s41588-022-01033-y

62. CDC COVID-19 Response Team. SARS-CoV-2 B.1.1.529 (Omicron) variant-United States, December 1-8, 2021. *MMWR Morb Mortal Wkly Rep.* 2021;70(50):1731-1734.

doi: 10.15585/mmwr.mm7050e1

63. Lambrou AS, Shirk P, Steele MK, *et al.* Genomic surveillance for SARS-CoV-2 variants: predominance of the delta (B.1.617.2) and Omicron (B.1.1.529) variants-United States, June 2021-January 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71(6):206-211.

doi: 10.15585/mmwr.mm7106a4

64. The European Centre for Disease Prevention and Control. *ECDC De-escalates BA.2, BA.4 and BA.5 from its List of Variants of Concern*; 2025. Available from: https://www.ecdc.europa.eu/en/news-events/ecdc-de-escalates-ba2-ba4-and-ba5-its-list-variants-concern [Last accessed on 2025 May 09].

65. Kanuri S, Chen A, Barnett D, Hsu E. CDC policy changes during the COVID-19 pandemic. *Open J Soc Sci*. 2024;12:174-200.

doi: 10.4236/jss.2024.126009

66. Dolan PT, Whitfield ZJ, Andino R. Mechanisms and concepts in RNA virus population dynamics and evolution. *Annu Rev Virol*. 2018;5(1):69-92.

doi: 10.1146/annurev-virology-101416-041718

67. Woo HJ, Reifman J. Quantitative modeling of virus evolutionary dynamics and adaptation in serial passages using empirically inferred fitness landscapes. *J Virol*. 2014;88(2):1039-1050.

doi: 10.1128/jvi.02958-13

68. Lancaster KZ, Pfeiffer JK. Viral population dynamics and virulence thresholds. *Curr Opin Microbiol*. 2012;15(4):525-530.

doi: 10.1016/j.mib.2012.05.007

69. Jin T, Yin J. Patterns of virus growth across the diversity of life. *Integr Biol (Camb)*. 2021;13(2):44-59.

doi: 10.1093/intbio/zyab001

70. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet*. 2009;10(8):540-550.

doi: 10.1038/nrg2583

71. Lv JX, Liu X, Pei YY, *et al.* Evolutionary trajectory of diverse SARS-CoV-2 variants at the beginning of COVID-19 outbreak. *Virus Evol.* 2024;10(1):veae020.

doi: 10.1093/ve/veae020

72. Grellet E, L'Hôte I, Goulet A, Imbert I. Replication of the coronavirus genome: A paradox among positive-strand

RNA viruses. *J Biol Chem*. 2022;298(5):101923.

doi: 10.1016/j.jbc.2022.101923

73. Gupta S, Gupta D, Bhatnagar S. Analysis of SARS-CoV-2 genome evolutionary patterns. *Microbiol Spectr*. 2024;12(2):e0265423.

doi: 10.1128/spectrum.02654-23

74. Fischer W, Giorgi EE, Chakraborty S, *et al*. HIV-1 and SARS-CoV-2: Patterns in the evolution of two pandemic pathogens. *Cell Host Microbe*. 2021;29(7):1093-1110.

doi: 10.1016/j.chom.2021.05.012

75. Wang X, Li J, Liu H, Hu X, Lin Z, Xiong N. SARS-CoV-2 versus influenza A virus: Characteristics and co-treatments. *Microorganisms*. 2023;11(3):580.

doi: 10.3390/microorganisms11030580

76. Tahir M. Coronavirus genomic nsp14-ExoN, structure, role, mechanism, and potential application as a drug target. *J Med Virol*. 2021;93(7):4258-4264.

doi: 10.1002/jmv.27009

77. Robson F, Khan KS, Le TK, *et al*. Coronavirus RNA proofreading: Molecular basis and therapeutic targeting. *Mol Cell*. 2020;79(5):710-727.

doi: 10.1016/j.molcel.2020.07.027

78. Simmonds P. Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: Causes and consequences for their short- and long-term evolutionary trajectories. *mSphere*. 2020;5(3):e00408-20.

doi: 10.1128/mSphere.00408-20

79. Neher RA. Contributions of adaptation and purifying selection to SARS-CoV-2 evolution. *Virus Evol*. 2022;8(2):veac113.

doi: 10.1093/ve/veac113

80. Liu J, Wu Y, Gao GF. A structural voyage toward the landscape of humoral and cellular immune escapes of SARS-CoV-2. *Immunol Rev*. 2025;330(1):e70000.

doi: 10.1111/imr.70000

81. Tong Y, Lavillette D, Li Q, Zhong J. Role of hepatitis C virus envelope glycoprotein E1 in virus entry and assembly. *Front Immunol*. 2018;9:1411.

doi: 10.3389/fimmu.2018.01411

82. Beretta M, Migraine J, Moreau A, *et al*. Common evolutionary features of the envelope glycoprotein of HIV-1 in patients belonging to a transmission chain. *Sci Rep*. 2020;10(1):16744.

doi: 10.1038/s41598-020-73975-4

83. Xu C, Zhang N, Yang Y, *et al*. Immune escape adaptive mutations in hemagglutinin are responsible for the antigenic drift of Eurasian avian-like H1N1 swine influenza viruses. *J Virol*. 2022;96(16):e0097122.

doi: 10.1128/jvi.00971-22

84. Kuiken C, Korber B, Shafer RW. HIV sequence databases. *AIDS Rev*. 2003;5(1):52-61.

85. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database. *Bioinformatics*. 2005;21(3):379-384.

doi: 10.1093/bioinformatics/bth485

86. Calhoun VC, Hatcher EL, Yankie L, Nawrocki EP. Influenza sequence validation and annotation using VADR. *Database (Oxford)*. 2024;2024:baae091.

doi: 10.1093/database/baae091

87. Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res*. 2015;43(Database issue):D571-D577.

doi: 10.1093/nar/gku1207

88. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88.

doi: 10.1371/journal.pbio.0040088

89. Abecasis A, Vandamme AM. Origin and distribution of HIV-1 subtypes. In: Hope TJ, Stevenson M, Richman D, editors. *Encyclopedia of AIDS*. Berlin: Springer New York; 2014. p. 1-16.

90. Hemelaar J, Elangovan R, Yun J, *et al*. Global and regional molecular epidemiology of HIV-1, 1990-2015: A systematic review, global survey, and trend analysis. *Lancet Infect Dis*. 2019;19(2):143-155.

doi: 10.1016/s1473-3099(18)30647-9

91. Smith DB, Bukh J, Kuiken C, *et al*. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology*. 2014;59(1):318-327.

doi: 10.1002/hep.26744

92. Petrova VN, Russell CA. The evolution of seasonal influenza viruses. *Nat Rev Microbiol*. 2018;16(1):47-60.

doi: 10.1038/nrmicro.2017.118

93. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.

doi: 10.1093/molbev/mst010

94. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725-2729.

doi: 10.1093/molbev/mst197

95. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method.

Proc Natl Acad Sci U S A. 2004;101(30):11030-11035.

doi: 10.1073/pnas.0404206101

96. Ren A, Ishida T, Akiyama Y. Assessing statistical reliability of phylogenetic trees via a speedy double bootstrap method. *Mol Phylogenet Evol.* 2013;67(2):429-435.

doi: 10.1016/j.ympev.2013.02.011

97. Rambaut A, Drummond A. *FigTree v1. 3.1 Institute of Evolutionary Biology*. Scotland: University of Edinburgh. 2010.

98. Lemoine F, Gascuel O. The Bayesian phylogenetic bootstrap and its application to short trees and branches. *Mol Biol Evol.* 2024;41(11):msae238.

doi: 10.1093/molbev/msae238

99. Candido DS, Claro IM, de Jesus JG, *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. 2020;369(6508):1255-1260.

doi: 10.1126/science.abd2161

100. Tegally H, Wilkinson E, Lessells RJ, *et al.* Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat Med*. 2021;27(3):440-446.

doi: 10.1038/s41591-021-01255-3

101. Tegally H, Wilkinson E, Giovanetti M, *et al.* Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592(7854):438-443.

doi: 10.1038/s41586-021-03402-9

102. Day T, Gandon S, Lion S, Otto SP. On the evolutionary epidemiology of SARS-CoV-2. *Curr Biol*. 2020;30(15):R849-R857.

doi: 10.1016/j.cub.2020.06.031

103. Faria NR, Mellan TA, Whittaker C, *et al.* Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021;372(6544):815-821.

doi: 10.1126/science.abh2644

104. Focosi D, Maggi F. Recombination in coronaviruses, with a focus on SARS-CoV-2. *Viruses*. 2022;14(6):1239.

doi: 10.3390/v14061239

105. Turakhia Y, Thornlow B, Hinrichs A, *et al.* Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature*. 2022;609(7929):994-997.

doi: 10.1038/s41586-022-05189-9

106. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol*. 2011;9(8):617-626.

doi: 10.1038/nrmicro2614

107. Xiao Y, Rouzine IM, Bianco S, *et al.* RNA recombination enhances adaptability and is required for virus spread and virulence. *Cell Host Microbe*. 2016;19(4):493-503.

doi: 10.1016/j.chom.2016.03.009

108. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun*. 2020;11(1):5986.

doi: 10.1038/s41467-020-19818-2

109. Davies NG, Abbott S, Barnard RC, *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. 2021;372(6538):eabg3055.

doi: 10.1126/science.abg3055

110. Gutierrez B, Márquez S, Prado-Vivar B, *et al.* Genomic epidemiology of SARS-CoV-2 transmission lineages in Ecuador. *Virus Evol.* 2021;7(2):veab051.

doi: 10.1093/ve/veab051

111. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *Aids*. 2017;31(9):1211-1222.

doi: 10.1097/qad.0000000000001470

112. Hedskog C, Parhy B, Chang S, *et al.* Identification of 19 novel hepatitis C virus subtypes-further expanding HCV classification. *Open Forum Infect Dis*. 2019;6(3):ofz076.

doi: 10.1093/ofid/ofz076

113. Takeshita M, Nishina N, Moriyama S, *et al.* Immune evasion and chronological decrease in titer of neutralizing antibody against SARS-CoV-2 and its variants of concerns in COVID-19 patients. *Clin Immunol*. 2022;238:108999.

doi: 10.1016/j.clim.2022.108999

114. Mahilkar S, Agrawal S, Chaudhary S, *et al.* SARS-CoV-2 variants: Impact on biological and clinical outcome. *Front Med (Lausanne)*. 2022;9:995960.

doi: 10.3389/fmed.2022.995960

115. Quintero AM, Eisner M, Sayegh R, *et al.* Differences in SARS-CoV-2 clinical manifestations and disease severity in children and adolescents by infecting variant. *Emerg Infect Dis*. 2022;28(11):2270-2280.

doi: 10.3201/eid2811.220577

116. Cojocaru C, Cojocaru E, Turcanu AM, Zaharia DC. Clinical challenges of SARS-CoV-2 variants (Review). *Exp Ther Med*. 2022;23(6):416.

doi: 10.3892/etm.2022.11343

117. Parczewski M, Scheibe K, Witak-Jędra M, Pynka M, Aksak-Wąs B, Urbańska A. Infection with HIV-1 subtype D adversely affects the live expectancy independently of antiretroviral drug use. *Infect Genet Evol.* 2021;90:104754.

doi: 10.1016/j.meegid.2021.104754

118. Baeten JM, Chohan B, Lavreys L, *et al.* HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J Infect*

Dis. 2007;195(8):1177-1180.

doi: 10.1086/512682

119. Palm AA, Esbjörnsson J, Månsson F, *et al.* Faster progression to AIDS and AIDS-related death among seroincident individuals infected with recombinant HIV-1 A3/CRF02_AG compared with sub-subtype A3. *J Infect Dis*. 2014;209(5):721-728.

doi: 10.1093/infdis/jit416

120. Dorp CHV, Goldberg EE, Hengartner N, Ke R, Romero-Severson EO. Estimating the strength of selection for new SARS-CoV-2 variants. *Nat Commun*. 2021;12(1):7239.

doi: 10.1038/s41467-021-27369-3

121. Tongo M, Harkins GW, Dorfman JR, *et al.* Unravelling the complicated evolutionary and dissemination history of HIV-1M subtype A lineages. *Virus Evol.* 2018;4(1):vey003.

doi: 10.1093/ve/vey003

122. Bedford T, Riley S, Barr IG, *et al*. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*. 2015;523(7559):217-220.

doi: 10.1038/nature14460

123. Kwon Y, Park J, An E, Jung S, Kweon K. Impact of omicron-variant SARS-CoV-2 infection on depression and anxiety: A community-based study in Korea. *Psychiatry Investig*. 2024;21(4):415-421.

doi: 10.30773/pi.2023.0323

124. Feng X, Yang C, Yang H, *et al*. Anxiety, depression, and somatic symptom disorders in health care workers at high altitude during the rapid spread of the SARS-CoV-2 Omicron variant: A prospective cohort study. *Front Psychiatry*. 2022;13:1018391.

doi: 10.3389/fpsyt.2022.1018391

125. Janik E, Niemcewicz M, Podogrocki M, Majsterek I, Bijak M. The emerging concern and interest SARS-CoV-2 variants. *Pathogens*. 2021;10(6):633.

doi: 10.3390/pathogens10060633

126. Hattab D, Amer MFA, Al-Alami ZM, Bakhtiar A. SARS-CoV-2 journey: From alpha variant to omicron and its sub-variants. *Infection*. 2024;52(3):767-786.

doi: 10.1007/s15010-024-02223-y

127. Andre M, Lau LS, Pokharel MD, *et al.* From alpha to omicron: How different variants of concern of the SARS-Coronavirus-2 impacted the world. *Biology (Basel)*. 2023;12(9):1267.

doi: 10.3390/biology12091267

128. Thakur S, Sasi S, Pillai SG, *et al.* SARS-CoV-2 mutations and their impact on diagnostics, therapeutics and vaccines. *Front Med (Lausanne)*. 2022;9:815389.

doi: 10.3389/fmed.2022.815389

129. Hirabara SM, Serdan TDA, Gorjao R, *et al.* SARS-COV-2 Variants: Differences and potential of immune evasion. *Front Cell Infect Microbiol*. 2021;11:781429.

doi: 10.3389/fcimb.2021.781429

130. Jabeen M, Shoukat S, Shireen H, Bao Y, Khan A, Abbasi AA. Unraveling the genetic variations underlying virulence disparities among SARS-CoV-2 strains across global regions: Insights from Pakistan. *Virol J*. 2024;21(1):55.

doi: 10.1186/s12985-024-02328-8

131. Telenti A, Hodcroft EB, Robertson DL. The evolution and biology of SARS-CoV-2 variants. *Cold Spring Harb Perspect Med*. 2022;12(5):a041390.

doi: 10.1101/cshperspect.a041390

132. Kim H, Webster RG, Webby RJ. Influenza virus: Dealing with a drifting and shifting pathogen. *Viral Immunol*. 2018;31(2):174-183.

doi: 10.1089/vim.2017.0141

133. Burrell CJ, Howard CR, Murphy FA. Chapter 2 - Classification of viruses and phylogenetic relationships. In: Burrell CJ, Howard CR, Murphy FA, editors. *Fenner and White's Medical Virology*. 5th ed. United States: Academic Press; 2017. p. 15-25.

134. Moya A, Holmes EC, González-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol*. 2004;2(4):279-288.

doi: 10.1038/nrmicro863

135. Geretti AM. HIV-1 subtypes: Epidemiology and significance for HIV management. *Curr Opin Infect Dis*. 2006;19(1):1-7.

doi: 10.1097/01.qco.0000200293.45532.68

136. Nastri BM, Pagliano P, Zannella C, *et al.* HIV and drug-resistant subtypes. *Microorganisms*. 2023;11(1):221.

doi: 10.3390/microorganisms11010221

137. Santoro MM, Perno CF. HIV-1 genetic variability and clinical implications. *ISRN Microbiol*. 2013;2013:481314.

doi: 10.1155/2013/481314

138. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet*. 2004;5(1):52-61.

doi: 10.1038/nrg1246

139. Echeverría N, Moratorio G, Cristina J, Moreno P. Hepatitis C virus genetic variability and evolution. *World J Hepatol*. 2015;7(6):831-845.

doi: 10.4254/wjh.v7.i6.831

140. Martinez MA, Franco S. Therapy implications of hepatitis C virus genetic diversity. *Viruses*. 2020;13(1):41.

doi: 10.3390/v13010041

141. Sallam M, Khalil R. Contemporary insights into hepatitis C

virus: A comprehensive review. *Microorganisms*. 2024;12(6):1035.

doi: 10.3390/microorganisms12061035

142. Le Guillou-Guillemette H, Vallet S, Gaudy-Graffin C, *et al.* Genetic diversity of the hepatitis C virus: Impact and issues in the antiviral therapy. *World J Gastroenterol*. 2007;13(17):2416-2426.

doi: 10.3748/wjg.v13.i17.2416

143. Wu NC, Wilson IA. Influenza hemagglutinin structures and antibody recognition. *Cold Spring Harb Perspect Med*. 2020;10(8):a038778.

doi: 10.1101/cshperspect.a038778

144. Wu NC, Wilson IA. A perspective on the structural and functional constraints for immune evasion: Insights from influenza virus. *J Mol Biol*. 2017;429(17):2694-2709.

doi: 10.1016/j.jmb.2017.06.015

145. Faraji N, Zeinali T, Joukar F, *et al.* Mutational dynamics of SARS-CoV-2: Impact on future COVID-19 vaccine strategies. *Heliyon*. 2024;10(9):e30208.

doi: 10.1016/j.heliyon.2024.e30208

146. Ma KC, Shirk P, Lambrou AS, *et al.* Genomic surveillance for SARS-CoV-2 variants: circulation of omicron lineages - United States, January 2022-May 2023. *MMWR Morb Mortal Wkly Rep.* 2023;72(24):651-656.

doi: 10.15585/mmwr.mm7224a2

147. Ma KC, Castro J, Lambrou AS, *et al.* Genomic surveillance for SARS-CoV-2 variants: Circulation of omicron XBB and JN.1 lineages - United States, May 2023-September 2024. *MMWR Morb Mortal Wkly Rep.* 2024;73(42):938-945.

doi: 10.15585/mmwr.mm7342a1

148. Grabowski F, Preibisch G, Giziński S, Kochańczyk M, Lipniacki T. SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations. *Viruses*. 2021;13(3):392.

doi: 10.3390/v13030392

149. Havasi A, Visan S, Cainap C, Cainap SS, Mihaila AA, Pop LA. Influenza A, influenza B, and SARS-CoV-2 similarities and differences - a focus on diagnosis. *Front Microbiol*. 2022;13:908525.

doi: 10.3389/fmicb.2022.908525

150. Gruell H, Vanshylla K, Korenkov M, *et al.* SARS-CoV-2 Omicron sublineages exhibit distinct antibody escape patterns. *Cell Host Microbe*. 2022;30(9):1231-1241.e6.

doi: 10.1016/j.chom.2022.07.002

151. Stefic K, Bouvin-Pley M, Braibant M, Barin F. Impact of HIV-1 diversity on its sensitivity to neutralization. *Vaccines (Basel)*. 2019;7(3):74.

doi: 10.3390/vaccines7030074

152. van de Sandt CE, Kreijtz JH, Rimmelzwaan GF. Evasion of influenza A viruses from innate and adaptive immune responses. *Viruses*. 2012;4(9):1438-1476.

doi: 10.3390/v4091438

153. Nuwarda RF, Alharbi AA, Kayser V. An overview of influenza viruses and vaccines. *Vaccines (Basel)*. 2021;9(9):1032.

doi: 10.3390/vaccines9091032

154. Santiago GA, Volkman HR, Flores B, *et al.* SARS-CoV-2 omicron replacement of delta as predominant variant, Puerto Rico. *Emerg Infect Dis*. 2023;29(4):855-857.

doi: 10.3201/eid2904.221700

155. Zhao H, Han K, Gao C, *et al.* VOC-alarm: Mutation-based prediction of SARS-CoV-2 variants of concern. *Bioinformatics*. 2022;38(14):3549-3556.

doi: 10.1093/bioinformatics/btac370

156. Salehi-Vaziri M, Fazlalipour M, Seyed Khorrami SM, *et al.* The ins and outs of SARS-CoV-2 variants of concern (VOCs). *Arch Virol*. 2022;167(2):327-344.

doi: 10.1007/s00705-022-05365-2

157. Tegally H, Wilkinson E, Tsui JL, *et al.* Dispersal patterns and influence of air travel during the global expansion of SARS-CoV-2 variants of concern. *Cell*. 2023;186(15):3277-3290.e16.

doi: 10.1016/j.cell.2023.06.001

158. Chen S, Guo L, Xie Y, *et al.* Government responses to the COVID-19 pandemic of the Gulf Cooperation Council countries: Good practices and lessons for future preparedness. *Glob Health Res Policy*. 2024;9(1):10.

doi: 10.1186/s41256-024-00349-y

159. Kisa S, Kisa A. A comprehensive analysis of COVID-19 misinformation, public health impacts, and communication strategies: Scoping review. *J Med Intern Res*. 2024;26:e56931.

doi: 10.2196/56931