

Article

Emergence of Genomic Diversity in the Spike Protein of the “Omicron” Variant

Surajit Basak ^{1,*} , Pratanu Kayet ¹, Manisha Ghosh ¹, Joyeeta Chatterjee ²  and Shanta Dutta ³

¹ Division of Bioinformatics, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata 700010, India

² Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry 605014, India

³ Division of Bacteriology, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata 700010, India

* Correspondence: basaksurajit@gmail.com; Tel.: +91-986-292-4152

Abstract: SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus) has constantly been evolving into different forms throughout its spread in the population. Emerging SARS-CoV-2 variants, predominantly the variants of concern (VOCs), could have an impact on the virus spread, pathogenicity, and diagnosis. The recently emerged “Omicron” variant has exhibited rapid transmission and divergence. The spike protein of SARS-CoV-2 has consistently been appearing as the mutational hotspot of all these VOCs. In order to determine a deeper understanding of the recently emerged and extremely divergent “Omicron”, a study of amino acid usage patterns and their substitution patterns was performed and compared with those of the other four successful variants of concern (“Alpha”, “Beta”, “Gamma”, and “Delta”). We observed that the amino acid usage of “Omicron” has a distinct pattern that distinguishes it from other VOCs and is significantly correlated with the increased hydrophobicity in spike proteins. We observed an increase in the non-synonymous substitution rate compared with the other four VOCs. Considering the phylogenetic relationship, we hypothesized about the functional interdependence between recombination and the mutation rate that might have resulted in a shift in the optimum of the mutation rate for the evolution of the “Omicron” variant. The results suggest that for improved disease prevention and control, more attention should be given to the significant genetic differentiation and diversity of newly emerging variants.

Keywords: SARS-CoV-2; VOC; amino acid usage; mutation rate; recombination; receptor binding domain



Citation: Basak, S.; Kayet, P.; Ghosh, M.; Chatterjee, J.; Dutta, S.

Emergence of Genomic Diversity in the Spike Protein of the “Omicron” Variant. *Viruses* **2023**, *15*, 2132.

<https://doi.org/10.3390/v15102132>

Academic Editor: Alexandre Hassanin

Received: 9 August 2023

Revised: 13 September 2023

Accepted: 20 September 2023

Published: 21 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) responsible for the Coronavirus disease 2019 (COVID-19) pandemic has had a distressing effect globally [1,2]. The number of mutations that accumulate as the genome replicates increases, leading to the variants being produced. The emergence of SARS-CoV-2 has also generated interest in the role of recombination in the evolution of the virus [3].

Distinguished SARS-CoV-2 variants were categorized by the World Health Organization (WHO) into three groups: variants of concern (VOCs), variants of interest (VOIs), and variants under monitoring (VUMs) in order to prioritize the surveillance of these variants [4,5]. Evolution of mutated lineages with higher infectivity or immune escape capacities were eventually categorized as VOCs based on their characteristics [4]. As of December 2021, WHO has identified five VOCs (“Alpha”, “Beta”, “Gamma”, “Delta”, and “Omicron”). Among them, the “Alpha” VOC (Pango lineage B.1.1.7) was first detected in the U.K. in 2020, the “Beta” VOC (Pango lineage B.1.351) was first detected in South Africa in 2020, the “Gamma” VOC (Pango lineage P.1) was first detected in 2020 in Brazil, the “Delta” VOC (Pango lineage B.1.617.2) was first detected in India in late 2020, and the “Omicron” VOC (Pango lineage B.1.1.529) was first detected in South Africa and Botswana in November 2021. The “Omicron” variant outcompeted other variants with a greater

number of mutations, specifically in the spike (S) protein, which has been linked to the high transmissibility and infectivity of this variant [6–8].

The evolution and adaptability of the viral genome to the host genome have been related to the global spread of SARS-CoV-2, which has shown significant drops in some places and sharp increases in others in terms of transmission rates. The variability has also been altered by variations in host immune systems, mutations, deletions, recombination, and genetic drift [9]. Studies have shown that the patterns of global variation are probably caused by adaptations at the nucleotide and amino acid regions as well as the variability found in structural proteins, notably spike proteins [10]. Multivariate analysis of the amino acid usage pattern will help to understand the worldwide heterogeneity of SARS-CoV-2, the evolution of its genome, and its adaptability to the host. Studies have also shown natural selection and mutational pressure as key factors influencing the variation in SARS-CoV-2 [3,11].

Genome-wide analysis of SARS-CoV-2 variants with the first emerged genome in Wuhan city in China in 2019 might suggest a relative association among viral strains [9,12]. With the prior knowledge of the accumulation of nucleotide substitution over time, the estimation of the most recent common ancestor (tMRCA) of the two strains is possible [13]. As the SARS-CoV-2 genome is almost entirely made up of protein-coding regions, it is crucial to distinguish between nonsynonymous and synonymous substitution rates [14,15]. The SARS-CoV-2 genome can actually be thought of as a collection of several “recombination blocks”, or areas between predicted breakpoints for recombination events [16]. For the current pandemic, recombination events in the evolutionary history of the spike protein are specifically important [17].

This study aims to comprehend the evolutionary pattern and mutational landscape of the five VOCs (“Alpha”, “Beta”, “Delta”, “Gamma”, and “Omicron”). First, we will identify the major trends in amino acid usage of spike proteins across five VOCs through multivariate analysis. We will also evaluate the evolutionary perspective, such as nonsynonymous and synonymous substitutions rates and rate of evolution, and then analyze the recombination pattern of the five VOCs. This study will be important to understand the mutational and evolutionary properties that are necessary for new therapeutic and vaccine development to combat the virus.

2. Materials and Methods

2.1. Sequence Retrieval

A total of 456,409 spike nucleotide sequences of SARS-CoV-2 were downloaded from the NCBI-Virus repository (<https://www.ncbi.nlm.nih.gov/sars-cov-2> accessed on 20 March 2022). Partial sequences and sequences containing ambiguous characters were excluded from our dataset [1,18]. NCBI provides information about each spike sequence regarding its association with a designated VOC (i.e., “Alpha”/“Beta”/“Gamma”/“Delta”/“Omicron”). Finally, there were 100,309, 427, 175,351, 9618, and 170,243 numbers of “Alpha”, “Beta”, “Delta”, “Gamma”, and “Omicron” variant sequences, respectively, in the downloaded data (Supplementary Table S1). 250 whole genome sequences covering five VOCs of SARS-CoV-2 were downloaded from NCBI for the recombination study. Angiotensin-converting enzyme 2 (ACE2) variants were retrieved from the Genome Aggregation Consortium Database (gnomAD) (<https://gnomad.broadinstitute.org/> accessed on 1 September 2023) (Supplementary Table S2).

2.2. Correspondence Analysis on Amino Acid Usage

To assess the variations in the amino acid usage of the spike protein, we performed a correspondence analysis (CoA). Major trends in variance in the dataset were revealed by placing the data along continuous axes. We employed correspondence analysis available in CodonW v1.4.2 software for the amino acid usage analysis of spike gene sequences [19–21]. Determination of the hydrophobicity of each spike gene sequence was conducted using the Kyte–Doolittle method present in the CodonW program [1,22].

2.3. Analysis of Evolutionary Selection

The ratio (ω) of the rate of non-synonymous substitutions per nonsynonymous site (K_a) to the rate of synonymous substitutions per synonymous site (K_s) reveals the influence of evolution on a gene segment. $\omega > 1$ indicates diversifying (positive) selection, whereas $\omega < 1$ signifies purifying (negative) selection. The evolutionary rates of genes (with reference to consensus sequence) were calculated using the Codeml program included in the PAML software package (ver. 4.5) [23–26] (<http://abacus.gene.ucl.ac.uk/software/paml.html> accessed on 1 September 2023) with runmode = −2 and CodonFreq = 1 [26]. Spike gene sequences were subjected to the analysis of synonymous and non-synonymous substitution rates with respect to the reference SARS-CoV-2 spike gene sequence. Statistical tests such as the t-test were conducted using the GraphPad (<https://www.graphpad.com/> accessed on 1 September 2023) web application.

2.4. Detection of Mutation Rate and TMRCA

BEAST (Bayesian Evolutionary Analysis Sampling Trees) is a software suite for phylogenetic analysis with an importance on time-scaled trees. The BEAST-1.10.4 software package was used to estimate the mutation rate and TMRCA (Time to Most Recent Common Ancestor) of the spike gene sequences [27]. Gene sequences were aligned using MAFFT v.7 [28] and the best-fit evolutionary model was estimated in MEGA-X [29]. The BEAUti2 v1.10.4 graphical user interface tool was used for generating the configuration files. The Tracer [30] and FigTree v1.4.4 tools were used for analyzing and visualizing the log data.

2.5. Recombination Analysis

A recombination analysis was performed with the whole genome sequences using the RDP4 program [31]. From the GISAID database, we have taken one whole genome from each lineage of all variants of concern (VOCs) of SARS-CoV-2 for our analysis and aligned them against the reference SARS-CoV-2 genome using MAFFT v7 [28]. The full-genome alignment was scanned for recombination using different algorithms. RDP, GENECONV, MaxChi, Chimera, and 3Seq algorithms were used for the primary scan. BootScan and SiScan algorithms were used for the secondary scan [32–34].

2.6. Protein Homology Modeling and Docking

Three-dimensional structural models of the spike protein of five VOCs were generated through homology modeling using the MODELER program [35]. Similarly, the structure of angiotensin-converting enzyme 2 (ACE2) was also generated. Protein structural models generated through homology modeling were refined using the ModRefiner web server [36]. The molecular interaction between viral spike proteins and the human ACE2 receptor was studied using a Z-dock server [37]. Then, using the PRODIGY webserver [38], the resulting docking data were processed and analyzed while taking the binding energies of each complex into consideration.

3. Results

3.1. Analysis of Amino Acid Usage

We performed a correspondence analysis to detect the amino acid usage pattern among the spike gene sequences of five SARS-CoV-2 variants. We observed a clear separation of the spike genes of the “Omicron” variant from the other four variants along the first major axis (Figure 1) that explained 82.86% of the total variations, while no other axis could explain more than 9.36% of the total variations.

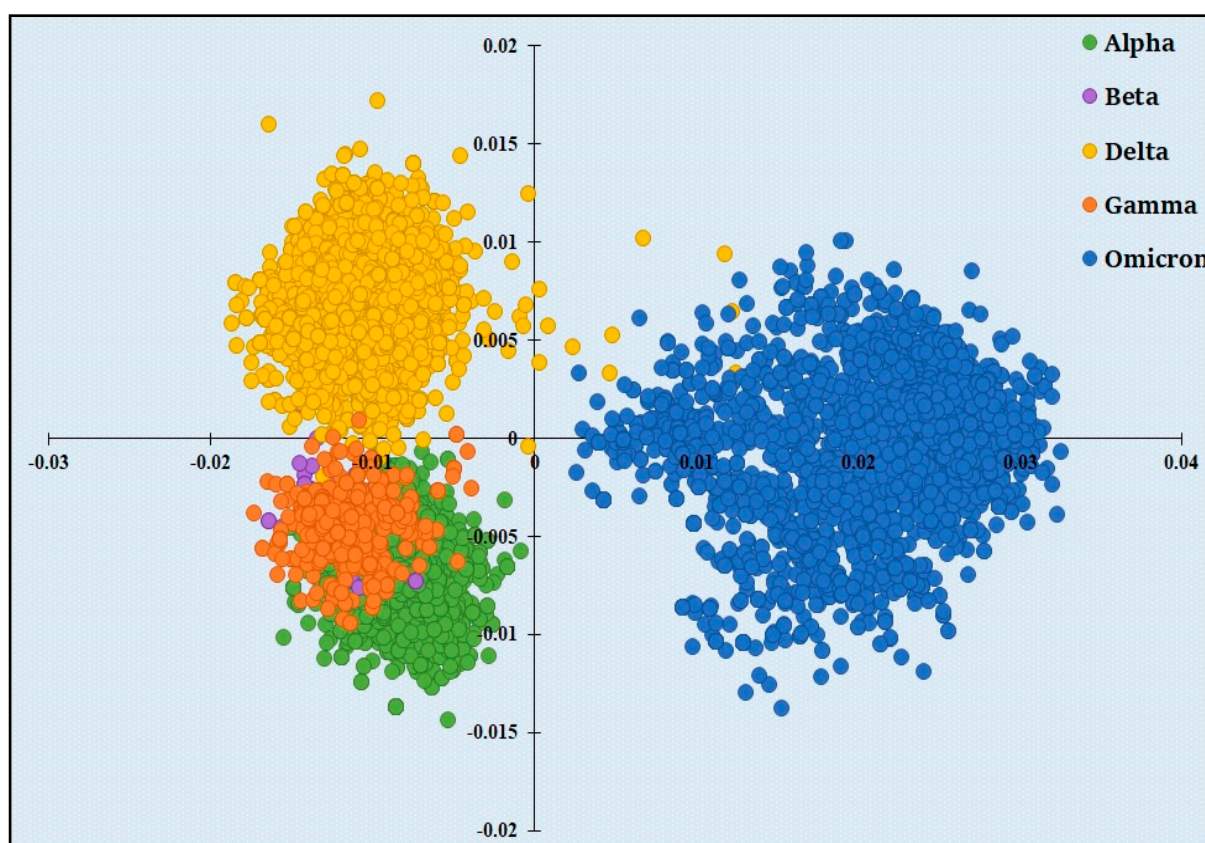


Figure 1. Distribution of spike genes of five VOCs along the two major axes of the correspondence analysis (COA) based on amino acid usage (AAU) data. *x*-axis—Axis 1 of AAU; *y*-axis—Axis 2 of AAU. Spike genes of “Alpha”, “Beta”, “Delta”, “Gamma”, and “Omicron” are represented with green, purple, yellow, orange, and blue dots, respectively.

We observed a positive correlation between the hydrophobicity of the encoded proteins and the position of the genes along the horizontal axis ($r = 0.238$, $p < 0.01$). Additionally, we found that the average hydrophobicity of the spike proteins distributed on the negative side (-0.0821) of the horizontal axis is significantly lower ($p < 0.01$) than the average hydrophobicity of the spike proteins distributed on the positive side (-0.0795) of the horizontal axis. These results clearly state that the evolution of hydrophobicity in spike proteins has been associated with the accumulation of more hydrophobic residues in the “Omicron” variant with respect to “Delta” and “Alpha” variants.

According to the amino acid composition, there is a rise in the following amino acid compositions of the “Omicron” variant compared with the “Delta” variant: Arginine, Lysine, Aspartic acid, and Glutamic acid. These increases indicate that the “Omicron” variant has more charged residues that contribute to salt bridge formation and that charged residues are exposed to a much greater degree.

The higher amino acid composition of Phenylalanine and Isoleucine in the “Omicron” spike protein, when compared with the “Delta” variant, suggests that the “Omicron” spike protein includes more hydrophobic amino acids, which may be due to its positioning inside the protein core. When compared with the “Delta” variant, the “Omicron” variant’s amino acid composition is low in polar amino acids such as Asparagine and Glutamine.

3.2. Evolutionary Rate Analysis

Estimation of synonymous (K_s) and nonsynonymous (K_a) substitution rates is important in understanding the dynamics of molecular sequence evolution. The K_a values for the spike gene sequences were found to correlate significantly with the data points on Axis 1 ($r = 0.98$, $p < 0.01$). The K_s values also correlated significantly with the data

points on Axis 1 ($r = 0.91$, $p < 0.01$). These results indicate that evolutionary selection pressure significantly influenced the amino acid usage pattern of spike gene sequences of five variants of SARS-CoV-2.

Our results show that the Ka/Ks values of the spike genes of the “Alpha” variant are the lowest and under purifying selection. Gradually, the values of Ka/Ks increase to reach the value closer to 1 (1.008) for the “Delta” variant, indicating that the genome was going through a neutral evolution (Figure 2). The values of Ka/Ks become much higher for the “Omicron” variant and under positive selection. This increase in Ka/Ks is mostly due to the enhanced rate of the nonsynonymous substitution rate; in particular, the nonsynonymous substitution rate is increased more than four times in the “Omicron” variant compared with the “Delta” variant. We hypothesize that this increase may be attributed to a larger diversity of sequences that may have given rise to more diverse lineages via undetected intra-SARS-CoV-2 recombination, which is analogous to a positive feedback loop.

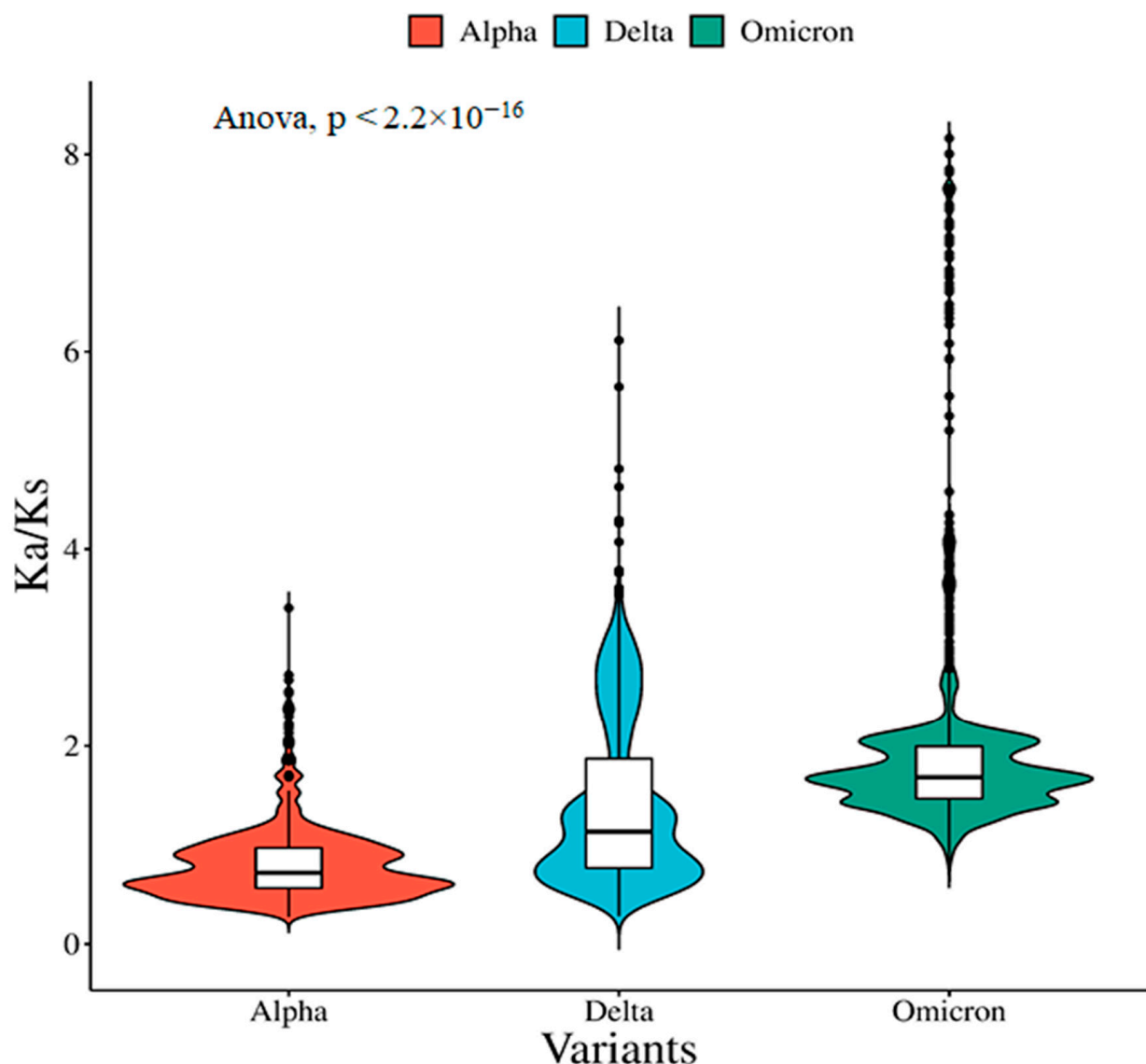


Figure 2. Distribution and statistical comparison of the Ka/Ks ratio of spike genes among the “Alpha”, “Delta”, and “Omicron” variants. The plot shows a significantly higher distribution of the Ka/Ks ratio of the “Omicron” variant compared with the “Alpha” and “Delta” variants.

The relatively high K_a/K_s (Figure 2) ratio for the “Omicron” variant suggests that the selective pressure acting on the spike protein of the “Omicron” variant is relaxed, and some sites may be undergoing positive selection. This increased evolutionary rate can be explained by the important function of the spike protein, which participates in host-specific recognition and undergoes several drastic changes during virus infection.

3.3. Interaction Profile between Spike Protein and ACE2

Our comparison of amino acid usage underlines the differential pattern of evolution through the accumulation of mutations in spike proteins among the five SARS-CoV-2 variants. Since the receptor for SARS-CoV-2 has been identified as ACE2, it was very important to analyze how the differential mutation patterns of amino acid usages of the spike proteins of five variants of SARS-CoV-2 responded to binding to the human ACE2 receptor. Three-dimensional structures of spike protein sequences for spike proteins from each of the five variants were constructed through homology modeling. The 3D structure of ACE2 was also generated computationally through homology modeling. The docking study was performed with ACE2 separately with five spike proteins representing five different VOCs and the binding energy was calculated separately for each of the docking experiments. We observed that the binding energy for the spike–ACE2 complex for the “Omicron” variant is lowest among the binding energy of the four other complexes (Figure 3). The “Omicron” variant always represented the lowest energy complex with all the ACE2 variants when the ACE2 variants were docked with the spike protein variants. A lower binding energy for the spike–ACE2 complex for the “Omicron” variant indicates its higher stability compared with the other four complexes made by the ACE2 and spike protein from the “Alpha”, “Beta”, “Gamma” and “Delta” variants. The highest stability of the spike (“Omicron”)-ACE2 complex may also be corroborated by the presence of more hydrogen bonds in the spike (“Omicron”)-ACE2 complex (Supplementary Figure S1).

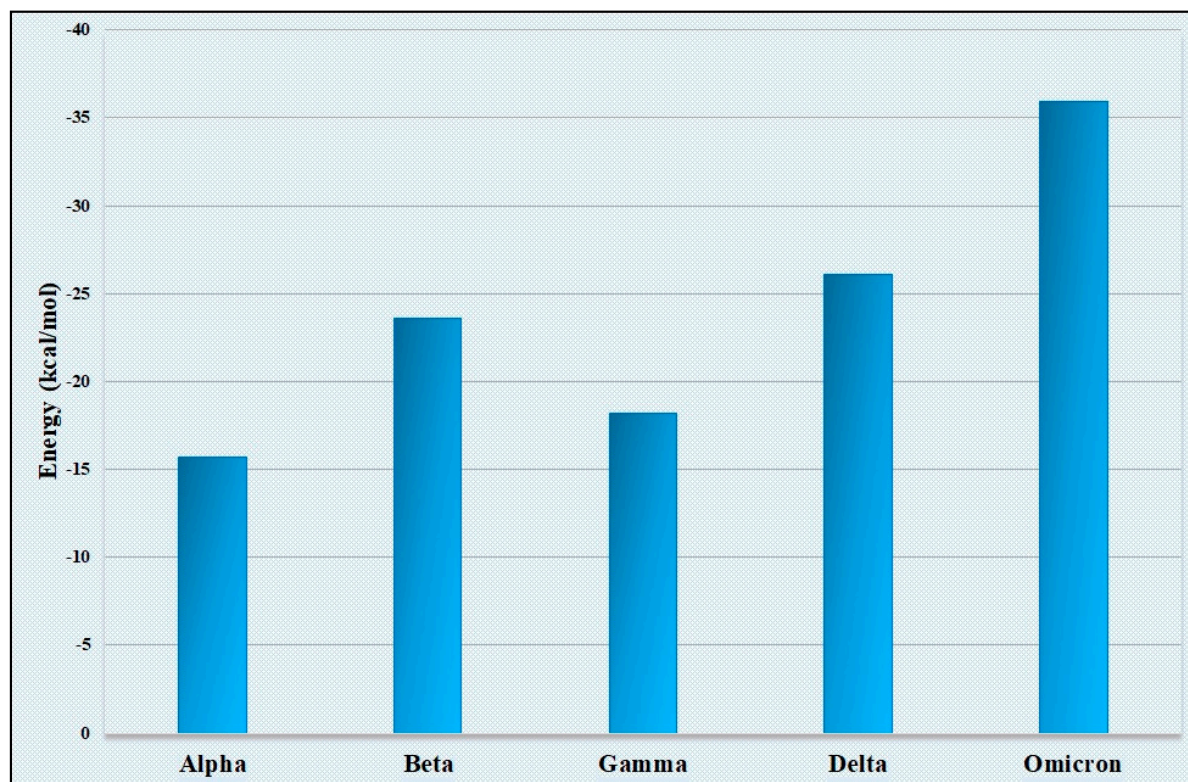


Figure 3. Variation in the binding energy of the spike–ACE2 complex among the five VOCs showing lowest binding energy (energy in negative scale) in the “Omicron” variant.

We surveyed the Genome Aggregation Consortium Database (gnomAD) (<https://gnomad.broadinstitute.org/> accessed on 1 September 2023) and found that human ACE2 is highly polymorphic, with single-nucleotide variants (SNVs) that result in missense mutations. These ACE2 could influence susceptibility to SARS-CoV-2 and potentially affect disease outcomes. The information on the association of each of these ACE2 alleles in various populations (e.g., East Asian, South Asian, European, African, Admixed American, etc.) was collected from the same database. The number of ACE2 genes associated with each population is shown in Table 1. For a given population, the allele frequency was calculated and the allele with the highest frequency is considered as the most common allele in the given population. The binding energies between the most common allele of a population and the spike protein of Omicron are provided in Table 2. It is clear that the highest binding energy is represented by four different populations (viz. American, Jewish, European, and South Asian). African and East Asian populations represent lower binding energies. Our results support epidemiological evidence that the African continent has an extremely low incidence and fatality rate compared with America [39]. We observed lower binding energy for the African population compared with the American population, which, in turn, indicates that the African population will be less prone to SARS-CoV-2 infection due to enhanced binding affinity.

Table 1. The number of ACE2 genes associated with each population.

Population	The Number of Associated ACE2 Genes
African/African American	45
Latino/Admixed American	41
Ashkenazi Jewish	3
East Asian	26
Europe	139
South Asian	43

Table 2. The binding energies between the most common allele of a population and spike protein of Omicron.

Population	Most Common Allele	Binding Energy (kcal/mol)
African/African American	rs147311723	−23.4
Latino/Admixed American	rs4646116	−24.2
Ashkenazi Jewish	rs41303171	−24.2
East Asian	rs191860450	−23.9
Europe	rs41303171	−24.2
South Asian	rs41303171	−24.2

3.4. Mutation Rate and tMRCA

In this study, tMRCA results were in accordance with the reported emergence of all VOCs (designated as per WHO guidelines). The “Beta” and “Gamma” variants appeared almost simultaneously, followed by the “Delta” variant and then the “Omicron” variant (Table 3). This observation in time scale phylogeny helps us to understand that, in most cases, recombinant “Delta” lineages were created by other “Delta” lineages because all the “Delta” lineages shared a common ancestor at a time point during the 2nd wave of COVID-19.

Table 3. TMRCA values depicting the emergence of the VOCs. A ten-time increase in mutation rate is observed in the “Omicron” variant compared with the “Alpha” variant.

Spike Variant	Mutation Rate	Mean Value (TMRCA)	Root Age
“Alpha”	3.537×10^{-3}	1.109	2019.943
“Beta”	3.18×10^{-3}	0.95	2020.03
“Gamma”	3.737×10^{-3}	1.07	2020.006
“Delta”	7.25×10^{-3}	1.189	2020.967
“Omicron”	3.506×10^{-2}	0.746	2021.414

Recombination events in the evolutionary history of the spike protein have particular significance for the current pandemic. The spike protein sequences are known to undergo frequent changes through recombination. The recombination and tMRCA results show us that some of the early recombinant lineages of a particular variant were created by intra-variant recombination, e.g., recombination between several lineages of “Delta” like AY.80, AY.86, etc. But later recombinants like BA.2 were the results of inter-variant recombination (Table 4). In the case of the “Omicron” variant, we observed a nearly 10-time increase in mutation rate compared with the “Alpha” variant (Figure 4). The enhanced mutation rate might have been possible due to the higher rate of nonsynonymous substitution as observed in this study. The “Delta” and “Omicron” variants created recombinants like BA.2 because they all share a common ancestor in time scale phylogeny.



Figure 4. Graphical representation of the mutation rate of five VOCs. A ten-time increase in mutation rate is observed in the “Omicron” variant compared with the “Alpha” variant.

Table 4. Recombination analysis demonstrated that several recombinant events occurred among variants highlighted in the table. All recombinants and their respective major and minor parents are listed below. All are significant at $p < 0.01$.

Recombinant	Major Parent	Minor Parent
"Delta" AY99.1	"Delta" AY.126	"Delta" AY.106
"Delta" AY.88	"Gamma" P.1.5	"Delta" AY.99.1
"Delta" AY.34.2	"Delta" AY.34.1.1	"Delta" AY.88
"Delta" AY.86	"Delta" AY.105	"Delta" AY.106
"Delta" AY.126	"Delta" AY.90	"Delta" AY.20
"Delta" AY.80	"Delta" AY.85	"Delta" AY.90
"Delta" AY.46.6.1	"Delta" AY.56	"Delta" AY.88
"Omicron" BA.2	"Omicron" BA.1.1	"Delta" AY.2.0
"Omicron" BA.1.1	"Delta" AY.55	"Delta" AY.39

4. Discussion

This study compared the amino acid usage patterns of spike proteins of five VOC lineages ("Alpha", "Beta", "Gamma", "Delta", and "Omicron"). The distinct amino acid usage pattern of the "Omicron" variant from other variants is clear in Figure 1. The selection pressure on the evolution of spike genes in "Omicron" is expected to affect the distinct amino acid usage pattern. The non-synonymous (Ka) to synonymous substitutions (Ks) ratio in protein-coding genes is commonly used to detect the selection pressure during gene evolution. A Ka/Ks ratio larger than 1 indicates positive selection, while a Ka/Ks ratio less than 1 indicates negative selection acting on protein-coding genes. Our results show that the Ka/Ks of the "Alpha" variant of SARS-CoV-2 is less than 1 indicating purifying selection. The "Delta" variant showed a value almost equal to 1 indicating neutral selection. However, the comparatively high Ka/Ks ratio of the spike protein of the "Omicron" variant indicates that the selective pressure acting on this variant is relaxed, and positive selection may be occurring in some sites. This accelerated evolutionary rate can be explained by the crucial role of the spike protein, which contributes to host-specific recognition and goes through a number of significant modifications during virus infection.

Furthermore, our results from mutation rate analysis suggest that the rate of mutation significantly increases from the "Delta" to the "Omicron" variant. Observations from evolutionary selection suggest that "Alpha" variants face relatively greater purifying selection pressure during the 1st wave of the pandemic. But this scenario was changed during the 2nd wave, which was caused by the "Delta" variant. The "Delta" variant and its lineages appear to have been subjected to nearly neutral selection pressure from the beginning to the end of the wave. Again, these scenarios were changed during the emergence of the "Omicron" variant, which is associated with positive selection pressure.

Thus, it is worth noting that the rate of substitution is an important driving force for evolutionary selection, and the variation in evolutionary selection pressure may be critical for various waves throughout the pandemic period. Purifying selection pressure is associated with genomes of the "Alpha" variant during the first wave of the pandemic whose duration was shorter than the other two waves. In the case of the "Omicron" variant, positive evolutionary selection helps to establish its lineages in the human population longer than the other two variants through stronger binding affinity with the ACE2 receptor.

The phylogenetic tree (Figure 5) among the variants displays two clades formed after the divergence from the root. One clade includes "Alpha" (blue), "Beta" (green), "Delta" (violet), and "Omicron" (pink), and the other clade includes "Gamma" (red).

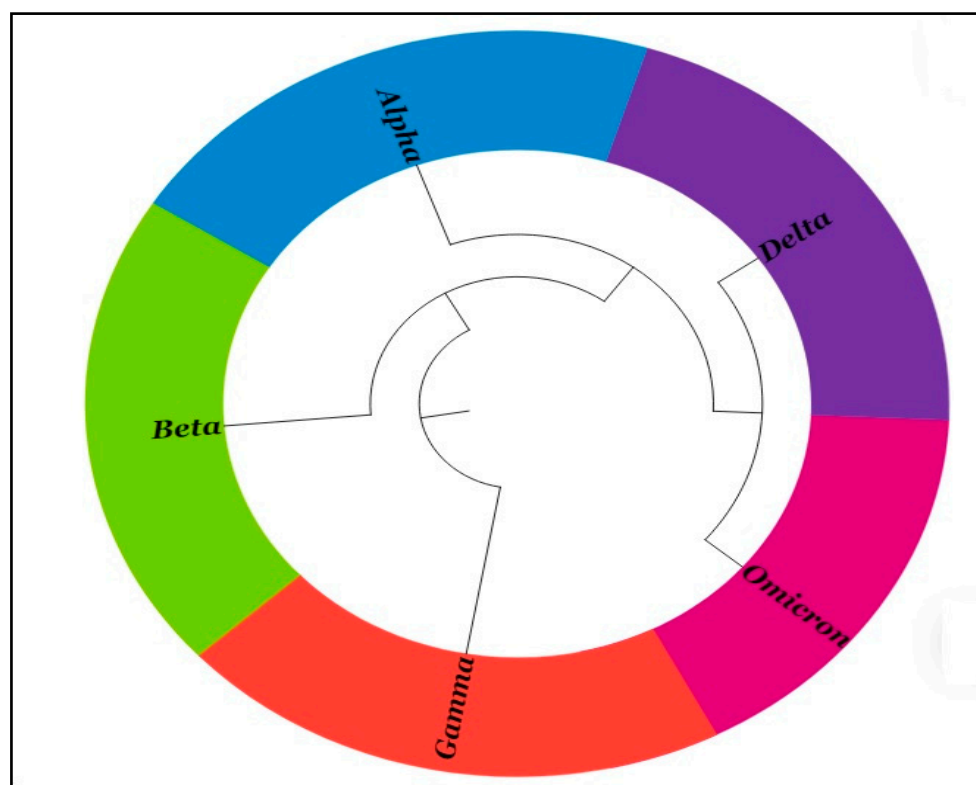


Figure 5. The phylogenetic tree among the variants shows that one group consists of “Alpha”, “Beta”, “Delta”, and “Omicron”, and another group consists of “Gamma”.

Our data show that the number of recombinants started increasing from the “Delta” variant, but its mutation rate was similar to other variants like “Alpha”, “Beta”, and “Gamma”. After that, a number of recombinants and a nearly 10-time increase in the mutation rate of the “Omicron” variant was observed. As an explanation for this occurrence, we prepared two hypotheses: one for the “Delta” variant, where we observed no shift in the mutation rate despite recombination. The other one is for the “Omicron” variant, where we observed a shift in mutation rate along with the recombination.

Hypothesis I: Recombination could accelerate adaptation without changing the optimal mutation rate. Earlier in the pandemic, diversity among SARS-CoV-2 was low and, eventually, the recombination rate was low. But later in the pandemic, various “Delta” lineages were created as a result of recombination. After this happened, recombination could speed up adaptation without changing the optimal mutation rate. This led to more differences between “Delta” lineages, but it did not change the optimal mutation rate.

Hypothesis II: “Omicron” was a product of the recombination of “Omicron” and “Delta” lineages. As a result, by the time of the emergence of this variant, genomic variation among the SARS-CoV-2 genomes was very high. Where there is a functional link between recombination and mutation, the rate of mutation shifts, and the ability of the host to adapt increases. Functional interdependence between recombination and the mutation rate would result in a shift in the optimum of the mutation rate.

The increase in the mutation rate in “Omicron” subvariants compared with other variants led to an increase in infectivity [40,41]. Although the accumulation of effective mutations is generally considered to be the main driving force behind viral evolution [42], inter-variant and intra-variant recombination can actually escalate this evolution process [43]. The high genetic diversity among the SARS-CoV-2 genomes during the time of evolution of the “Omicron” variant might be an important reason behind the large diversity observed among “Omicron” subvariants. The co-occurrence of multiple subvariants at a

similar time has made it possible for “Omicron” to undergo multiple recombinations due to co-infection of the host by different variants [44].

The increased recombination event in the “Omicron” variant has led to changes in the binding affinity between the Receptor Binding Domain (RBD) region of the spike protein and the human receptor ACE2. The increase in recombination rate will result in an increase in mutation rate in order to better the adaption capability of the virus. The high mutation rate in the “Omicron” variant increased the probability of accumulation of effective mutation in the spike proteins which serve as the key region in the viral genome to boost viral adaptation to the host. “Omicron” subvariants have a much higher rate of effective mutation among the other SARS-CoV-2 variants, which characterizes the viral adaptability through better binding of spike proteins to human receptors, thus increasing its infectivity.

5. Conclusions

The process of viral evolution is persistent and has the potential to enhance “viral fitness” and selective adaption. Scientists and authorities all around the world have suffered as a result of newly emerging SARS-CoV-2 strains. Vaccines presently offer good protection against all VOCs, but continuous monitoring of vaccine effectiveness is necessary to combat the main SARS-CoV-2 strains and the newly emerging variants. The advent of the “Omicron” variant and the evolution of the entire coronavirus subfamily serve as a warning to researchers, scientists, vaccine developers, and policymakers to maintain vigilance. However, the majority of currently approved vaccination plans and monoclonal antibody treatments target the spike ORFs because of their inherently higher rate of mutation and recombination; however, other more stable genomic regions should be thoroughly explored as potential targets for future research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15102132/s1>. Table S1: Accession numbers of Spike genes of five VOCs. Table S2: Information of ACE2 variants. Figure S1: Interaction profile of Spike protein and ACE2.

Author Contributions: Conceptualization, S.B, P.K., M.G., J.C. and S.D.; methodology, S.B.; software, S.B, P.K., M.G. and J.C.; validation, S.B, P.K., M.G. and J.C.; formal analysis, S.B, P.K., M.G., J.C. and S.D.; investigation, S.B and P.K.; resources, S.B.; data curation, S.B. and P.K.; writing—original draft preparation, S.B. and P.K.; writing—review and editing, S.B, P.K., M.G., J.C. and S.D.; visualization, S.B, P.K., M.G., J.C. and S.D.; supervision, S.B.; project administration, S.B.; funding acquisition, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Indian Council of Medical Research: 2021-3712.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Information of sequence data are available in the Tables S1 and S2. All original and processed data for the results of the present study are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ghosh, M.; Basak, S.; Dutta, S. Underlying selection for the diversity of spike protein sequences of SARS-CoV-2. *IUBMB Life* **2022**, *74*, 213–220. [CrossRef] [PubMed]
2. Lai, C.C.; Shih, T.P.; Ko, W.C.; Tang, H.J.; Hsueh, P.R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int. J. Antimicrob. Agents* **2020**, *55*, 105924. [CrossRef] [PubMed]
3. Khattak, S.; Rauf, M.A.; Zaman, Q.; Ali, Y.; Fatima, S.; Muhammad, P.; Li, T.; Khan, H.A.; Khan, A.A.; Ngowi, E.E.; et al. Genome-Wide Analysis of Codon Usage Patterns of SARS-CoV-2 Virus Reveals Global Heterogeneity of COVID-19. *Biomolecules* **2021**, *11*, 912. [CrossRef] [PubMed]

4. Vo, G.V.; Bagyinszky, E.; An, S.S.A. COVID-19 Genetic Variants and Their Potential Impact in Vaccine Development. *Microorganisms* **2022**, *10*, 598. [CrossRef] [PubMed]
5. Choi, J.Y.; Smith, D.M. SARS-CoV-2 Variants of Concern. *Yonsei Med. J.* **2021**, *62*, 961–968. [CrossRef] [PubMed]
6. Nikolaidis, M.; Papakyriakou, A.; Chlichlia, K.; Markoulatos, P.; Oliver, S.G.; Amoutzias, G.D. Comparative Analysis of SARS-CoV-2 Variants of Concern, Including Omicron, Highlights Their Common and Distinctive Amino Acid Substitution Patterns, Especially at the Spike ORF. *Viruses* **2022**, *14*, 707. [CrossRef] [PubMed]
7. Chakraborty, C.; Sharma, A.R.; Bhattacharya, M.; Mallik, B.; Nandi, S.S.; Lee, S.S. Comparative genomics, evolutionary epidemiology, and RBD-hACE2 receptor binding pattern in B.1.1.7 (Alpha) and B.1.617.2 (Delta) related to their pandemic response in UK and India. *Infect. Genet. Evol.* **2022**, *101*, 105282. [CrossRef] [PubMed]
8. Weng, S.; Shang, J.; Cheng, Y.; Zhou, H.; Ji, C.; Yang, R.; Wu, A. Genetic differentiation and diversity of SARS-CoV-2 Omicron variant in its early outbreak. *Biosaf. Health* **2022**, *4*, 171–178. [CrossRef]
9. Rahman, M.S.; Hoque, M.N.; Islam, M.R.; Akter, S.; Rubayet, A.S.M.; Siddique, M.A.; Saha, O.; Rahaman, M.M.; Sultana, M.; Crandall, K.A.; et al. Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: An in silico approach. *PeerJ* **2020**, *8*, e9572. [CrossRef]
10. Noureddine, F.Y.; Chakkour, M.; El Roz, A.; Reda, J.; Al Sahily, R.; Assi, A.; Joma, M.; Salami, H.; Hashem, S.J.; Harb, B.; et al. The Emergence of SARS-CoV-2 Variant(s) and Its Impact on the Prevalence of COVID-19 Cases in the Nabatieh Region, Lebanon. *Med. Sci.* **2021**, *9*, 40. [CrossRef]
11. Shen, Z.; Xiao, Y.; Kang, L.; Ma, W.; Shi, L.; Zhang, L.; Zhou, Z.; Yang, J.; Zhong, J.; Yang, D.; et al. Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients with Coronavirus Disease 2019. *Clin. Infect. Dis.* **2020**, *71*, 713–720. [CrossRef] [PubMed]
12. Kannan, S.R.; Spratt, A.N.; Sharma, K.; Chand, H.S.; Byrareddy, S.N.; Singh, K. Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. *J. Autoimmun.* **2022**, *126*, 102779. [CrossRef] [PubMed]
13. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef] [PubMed]
14. Zhao, Z.; Li, H.; Wu, X.; Zhong, Y.; Zhang, K.; Zhang, Y.P.; Boerwinkle, E.; Fu, Y.X. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* **2004**, *4*, 21. [CrossRef] [PubMed]
15. Singh, D.; Yi, S.V. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* **2021**, *53*, 537–547. [CrossRef] [PubMed]
16. Boni, M.F.; Lemey, P.; Jiang, X.; Lam, T.T.; Perry, B.W.; Castoe, T.A.; Rambaut, A.; Robertson, D.L. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* **2020**, *5*, 1408–1417. [CrossRef] [PubMed]
17. Huang, Y.; Yang, C.; Xu, X.F.; Xu, W.; Liu, S.W. Structural and functional properties of SARS-CoV-2 spike protein: Potential antiviral drug development for COVID-19. *Acta Pharmacol.* **2020**, *41*, 1141–1149. [CrossRef] [PubMed]
18. Roy, A.; Basak, S. HIV long-term non-progressors share similar features with simian immunodeficiency virus infection of chimpanzees. *J. Biomol. Struct. Dyn.* **2021**, *39*, 2447–2454. [CrossRef] [PubMed]
19. Peden, J.F. Analysis of Codon Usage. Nottingham: University of Nottingham. 2000. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=450270dddf16353d879274211d08b9fa7f1ea5537> (accessed on 15 June 2022).
20. Roy, A.; Banerjee, R.; Basak, S. HIV Progression Depends on Codon and Amino Acid Usage Profile of Envelope Protein and Associated Host-Genetic Influence. *Front. Microbiol.* **2017**, *8*, 1083. [CrossRef] [PubMed]
21. Ghosh, M.; Basak, S.; Dutta, S. Natural selection shaped the evolution of amino acid usage in mammalian toll like receptor genes. *Comput. Biol. Chem.* **2022**, *97*, 107637. [CrossRef]
22. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [CrossRef] [PubMed]
23. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **1986**, *3*, 418–426. [PubMed]
24. Sinha, N.K.; Roy, A.; Das, B.; Das, S.; Basak, S. Evolutionary complexities of swine flu H1N1 gene sequences of 2009. *Biochem. Biophys. Res. Commun.* **2009**, *390*, 349–351. [CrossRef] [PubMed]
25. Banerjee, R.; Roy, A.; Ahmad, F.; Das, S.; Basak, S. Evolutionary patterning of hemagglutinin gene sequence of 2009 H1N1 pandemic. *J. Biomol. Struct. Dyn.* **2012**, *29*, 733–742. [CrossRef] [PubMed]
26. Hurst, L.D. The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet.* **2002**, *18*, 486. [CrossRef] [PubMed]
27. Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **2007**, *7*, 214. [CrossRef] [PubMed]
28. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
29. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
30. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef]
31. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef]

32. Posada, D.; Crandall, K.A. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13757–13762. [[CrossRef](#)] [[PubMed](#)]
33. Lam, H.M.; Ratmann, O.; Boni, M.F. Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Mol. Biol. Evol.* **2018**, *35*, 247–251. [[CrossRef](#)] [[PubMed](#)]
34. Martin, D.P.; Posada, D.; Crandall, K.A.; Williamson, C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retrovir.* **2005**, *21*, 98–102. [[CrossRef](#)] [[PubMed](#)]
35. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5.6.1–5.6.37. [[CrossRef](#)] [[PubMed](#)]
36. Xu, D.; Zhang, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* **2011**, *101*, 2525–2534. [[CrossRef](#)] [[PubMed](#)]
37. Pierce, B.G.; Wiehe, K.; Hwang, H.; Kim, B.H.; Vreven, T.; Weng, Z. ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **2014**, *30*, 1771–1773. [[CrossRef](#)] [[PubMed](#)]
38. Honorato, R.V.; Koukos, P.I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A.M.J.J. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021**, *8*, 729513. [[CrossRef](#)] [[PubMed](#)]
39. Simeon, J.O.; Tosim, J.O.; Zubairu, S.A. Cumulative evaluation of demography and distribution of COVID-19 around the globe: An update report of COVID-19 until 17th February 2022. *Int. J. Epidemiol. Health Sci.* **2022**, *3*, e34.
40. Dawood, A.A. Increasing the frequency of Omicron variant mutations boosts the immune response and may reduce the virus virulence. *Microb. Pathog.* **2022**, *164*, 105400. [[CrossRef](#)]
41. Chakkour, M.; Salami, A.; Olleik, D.; Kamal, I.; Nouredine, F.Y.; Roz, A.E.; Ghssein, G. Risk Markers of COVID-19, a Study from South-Lebanon. *COVID* **2022**, *2*, 867–876. [[CrossRef](#)]
42. Kudriavtsev, A.V.; Vakhrusheva, A.V.; Novoseletsky, V.N.; Bozdaganyan, M.E.; Shaitan, K.V.; Kirpichnikov, M.P.; Sokolova, O.S. Immune Escape Associated with RBD Omicron Mutations and SARS-CoV-2 Evolution Dynamics. *Viruses* **2022**, *14*, 1603. [[CrossRef](#)]
43. Wang, L.; Zhou, H.Y.; Li, J.Y.; Cheng, Y.X.; Zhang, S.; Aliyari, S.; Wu, A.; Cheng, G. Potential intervariant and intravariant recombination of Delta and Omicron variants. *J. Med. Virol.* **2022**, *94*, 4830–4838. [[CrossRef](#)]
44. Rockett, R.J.; Draper, J.; Gall, M.; Sim, E.M.; Arnott, A.; Agius, J.E.; Johnson-Mackinnon, J.; Fong, W.; Martinez, E.; Drew, A.P.; et al. Co-infection with SARS-CoV-2 Omicron and Delta variants revealed by genomic surveillance. *Nat. Commun.* **2022**, *13*, 2745. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.