



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Original Article

# A comprehensive genomic study, mutation screening, phylogenetic and statistical analysis of SARS-CoV-2 and its variant omicron among different countries



Syed Umair Ahmad<sup>a</sup>, Bushra Hafeez Kiani<sup>b</sup>, Muhammad Abrar<sup>c</sup>, Zainab Jan<sup>a</sup>, Imran Zafar<sup>d</sup>, Yasir Ali<sup>e</sup>, Amer M. Alanazi<sup>f</sup>, Abdul Malik<sup>g</sup>, Mohd Ashraf Rather<sup>h</sup>, Asrar Ahmad<sup>i</sup>, Azmat Ali Khan<sup>f,\*</sup>

<sup>a</sup> Department of Bioinformatics, Hazara University, Mansehra, Pakistan

<sup>b</sup> Department of Biological Sciences, Faculty of Basic and Applied Sciences, International Islamic University Islamabad, 44000, Pakistan

<sup>c</sup> Department of Anesthesia, DHQ Teaching Hospital, Sahiwal Medical College, Sahiwal, Pakistan

<sup>d</sup> Department of Bioinformatics and Computational Biology, Virtual University, Pakistan

<sup>e</sup> National Centre for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan

<sup>f</sup> Pharmaceutical Biotechnology Laboratory, Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, P.O. Box 2457, Riyadh 11451, Saudi Arabia

<sup>g</sup> Department of Pharmaceutics, College of Pharmacy, King Saud University, P.O. Box 2457, Riyadh 11451, Saudi Arabia

<sup>h</sup> Division of Fish Genetics and Biotechnology, Faculty of Fisheries Ganderbal, Sher-e, Kashmir University of Agricultural Science and Technology, Kashmir, India

<sup>i</sup> Center for Sickle Cell Disease, College of Medicine, Howard University, Washington DC, USA

## ARTICLE INFO

## Article history:

Received 21 February 2022

Received in revised form 16 June 2022

Accepted 3 July 2022

## Keywords:

Viral genome

Mutation analysis

Phylogeny

Delta

Omicron variants

SARS-CoV-2

## ABSTRACT

**Background:** With the rapid development of the genomic sequence data for the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and its variants Delta (B.1.617.2) and Omicron (B.1.1.529), it is vital to successfully identify mutations within the genome.

**Objective:** The main objective of the study is to investigate the full-length genome mutation analysis of 157 SARS-CoV-2 and its variant Delta and Omicron isolates. This study also provides possible effects at the structural level to understand the role of mutations and new insights into the evolution of COVID-19 and evaluates the differential level analysis in viral genome sequence among different nations. We have also tried to offer a mutation snapshot for these differences that could help in vaccine formulation. This study utilizes a unique and efficient method of targeting the stable genes for the drug discovery approach.

**Methods:** Complete genome sequence information of SARS-CoV-2, Delta, and Omicron from online resources were used to predict structure domain identification, data mining, and screening; employing different bioinformatics tools. BioEdit software was used to perform their genomic alignments across countries and a phylogenetic tree as per the confidence of 500 bootstrapping values was constructed. Heterozygosity ratios were determined *in-silico*. A minimum spanning network (MSN) of selected populations was determined by Bruvo's distance role-based framework.

**Results:** Out of all 157 different strains of SARS-CoV-2 and its variants, and their complete genome sequences from different countries, Corona nucleocapsid and DUF5515 were observed to be the most conserved domains. All genomes obtained changes in comparison to the Wuhan-Hu-1 strain, mainly in the TRS region (CUAAAC or ACGAAC). We discovered 596 mutations in all genes, with the highest number (321) found in ORF1ab (QHD43415.1), or TRS site mutations found only in ORF7a (1) and ORF10 (2). The Omicron variant has 30 mutations in the Spike protein and has a higher alpha-helix shape (23.46%) than the Delta version (22.03%). T478 was also discovered to be a prevalent polymorphism in Delta and Omicron variations, as well as genomic gaps ranging from 45 to 65aa. All 157 sequences contained variations and conformed to Nei's Genetic distance. We discovered heterozygosity (Hs) 0.01, mean anticipated Hs 0.32, the genetic diversity index (GDI) 0.01943989, and GD within population 0.01266951. The Hedrick value was 0.52324978, the GD coefficient was 0.52324978, the average Hs was 0.01371452, and the GD coefficient was 0.52324978. Among

\* Corresponding author.

E-mail address: [azkhan@ksu.edu.sa](mailto:azkhan@ksu.edu.sa) (A.A. Khan).

other countries, Brazil has the highest standard error (SE) rate (1.398), whereas Japan has the highest ratio of Nei's gene diversity (0.01).

**Conclusions:** The study's findings will assist in comprehending the shape and kind of complete genome, their streaming genomic sequences, and mutations in various additions of SARS-CoV-2, as well as its different variant strains like Omicron. These results will provide a scientific basis to design the vaccines and understand the genomic study of these viruses.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of King Saud Bin Abdulaziz University for Health Sciences.

CC\_BY\_NC\_ND\_4.0

## Introduction

In December 2019, a disease pandemic that caused breathlessness in Wuhan, China, turned into a global epidemic [1,2]. The enterohaemorrhagic agents of COVID-19, a transmissible disease that has now spread worldwide in a matter of months, were discovered as the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [3]. Because of the introduction of new SARS-CoV-2 strains, the COVID-19 pandemic remains one of the primary causes of infectious death globally. The World Health Organization (WHO) officially declared the situation of worldwide significance on January 30, 2020 [4]. By January 6, 2022, 298 million instances had been documented and 5.47 million people had died in 221 countries around the world. With 1.3 million illnesses reported and 28,955 fatalities over the same period.

SARS-CoV-2 is a zoonotic virus that belongs to the beta coronavirus family and is extremely pathogenic [3]. This virus is dangerous due to three interconnected features: 1) because humans have no direct immunological experience with this virus, leaving us defenseless to infection and disease [5]; 2) it is exceptionally transmissible, and 3) it has a high mortality rate. Like other coronaviruses, SARS-CoV-2 is spherical, with a diameter of 125 nm and a genome size of ~30 kb. According to the genomic analysis, the virus appears to be a member of the beta-coronavirus 2b family [6]. The SARS-CoV-2 genome is made up of 29,903 positive polarities single-stranded RNA nucleotides [7]. Since it has a methylation cap at the 5' end and a polyadenylated tail (poly-A) at the 3' end, this RNA chain is highly homologous to messenger RNA (mRNA) of eukaryotic living cells [8]. Moreover, this viral genome has open reading frames (ORFs) coding sequences [9]. It has six ORFs that encode 16 non-structural proteins, as well as ORFs towards the 3' end of the genome [10]. It also encodes four main structural proteins: spike protein (S), membrane protein (M), an envelope protein (E), and nucleocapsid (N). The spike (S) is a glycoprotein that interacts with the cellular receptor angiotensin-converting enzyme 2 (ACE2) and is responsible for viral infection [11,12]. The S precursor protein of SARS-CoV-2 can be proteolytically cleaved into the S1 (685 amino acids) and S2 (588 amino acids) subunits [13].

The global & uncontrollable magnitude of such an outbreak created a circumstance in which genetic epidemiological was made a reality on a global scale, from the rapid sequencing of SARS-CoV-2 to the detection of potential lineages by continuous surveillance worldwide [14]. Before the COVID-19 epidemic, genomics data on the transmission of diseases were either scarce or non-existent. This set of circumstances was gradually altered with the emergence of SARS-CoV-2, but known evidence remains sparse [15]. Instead of active surveillance strategies or initiatives, academic researchers use transcriptome epidemiology studies to collect epigenetic data on SARS-CoV-2. This emphasizes the necessity of promoting regulatory frameworks that allow the use of genetic epidemiology as a strategy for improving monitoring and detection systems for potential health concerns in communities.

The mechanisms underlying the wide range of therapeutic vulnerability to COVID-19 and illness manifestation are not fully understood, raising concerns about the function of viral replication

genetic changes as potential factors influencing pathogenicity and autoimmune disease in particular. Understanding the pathogenesis of SARS-CoV-2 and laying the groundwork for the development of new vaccines and anti-retroviral therapies requires viral replication genomic research. Big data and artificial intelligence are two computational tools and methodologies that can help tickle SARS-CoV-2 and its subtypes Delta (B.1.617.2) and Omicron (B.1.1.529) [16]. SARS-CoV-2 with the Omicron mutant form was found in samples collected on November 11, 2021, in Botswana [17,18], as well as on November 14, 2021, in African countries. The first documented case of Omicron in the United States was seen on December 1, 2021, in California [19].

A subvariant of the highly transmissible Omicron variant called BA.2 appears to be taking over in certain countries. Omicron subvariant BA.2 could be 1.5 times more infectious than its predecessor – although they've detected no difference in the risk of hospitalization. As of 21 January, 2022, 40 countries had similarly uploaded BA.2 genome sequences on viral genome databases, with Denmark having recorded the most, followed by India, Sweden, and Singapore – although this could partly be due to higher levels of viral surveillance in these countries. Based on viral genomics we investigate the results of Omicron (BA. 2) SARS-CoV-2 Variant is more virulent and dangerous. When BA.2 rose to prominence, scientists were worried that its unique mutations might allow it to evade immunity from BA.1 infection. That doesn't seem to be the case. The World Health Organization said that infection with BA.1 provides strong protection against infection with BA.2. In countries like the United States that have come through major spikes of BA.1 infection, the immunity that has built up may shield them from a major BA.2 spike.

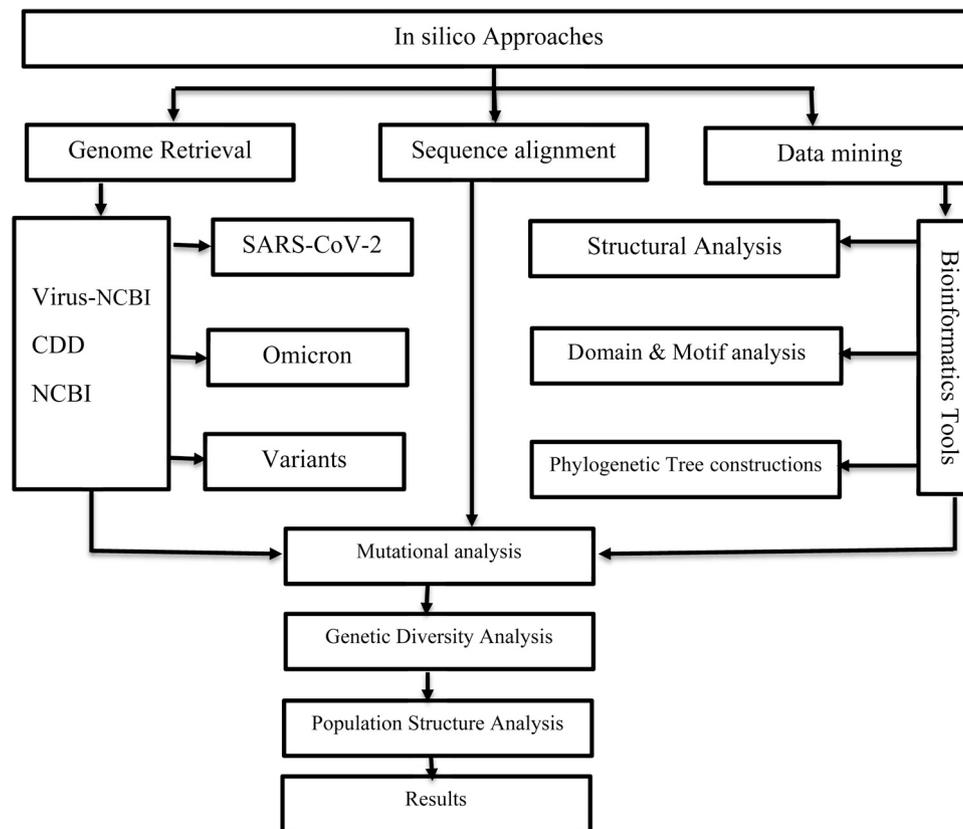
The goal of this work was to look at comparative genomics (variation and homology) of the SARS-CoV-2 virus genome from different countries, with an emphasis on identifying diversity by constructing an evolutionary phylogenetic tree and performing domain analysis worldwide. This study also focused on mutation screening and population based analysis.

## Material and methods

**Scheme 1** depicts the methodology used in this *in-silico* study. The flow chart depicts the workplan used for mutation screening and phylogenetic analysis of novel SARS-CoV-2 and its variants among different countries.

### Data retrieval

Uniport (AC# P0DTC2) provides the FASTA sequence of the Spike protein of SARS-CoV-2 wild type. The Delta variation of spike protein (AC# QWK65230.1) was investigated using the online database ViPR (<https://www.viprbrc.org/brc/home.spg?decorator=vipr>). The Omicron (R40B60) genome was retrieved from GSAID (<https://www.gisaid.org/>) and converted to protein sequence using the expasy translate program (<https://www.expasy.org/>). Based on the translated sequence, the Omicron spike protein was chosen. A systematic search was carried out in the GIRI database (Genetic



**Scheme 1.** Complete workflow of the *in-silico* strategy employed for screening of mutation and phylogenetic analysis of novel SARS-CoV-2 among different countries.

Information Research Institute ([girinst.org](http://girinst.org)) and virus NCBI (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>) to retrieve the 157 genomes of SARS-CoV-2 sequence from different countries.

#### Structure domain identification, data mining, and screening

The Omicron sequence was aligned with Delta and Omicron isoform variations using bioinformatics software. The alignment illustration was created using the box shading tool. For the identification of conserved domains in intact elements, the nucleotide sequences were considered in the conserved domain database (CDD) of NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) with default parameters. Conventional data mining approaches were used in CDD to retrieve conserved domains from the SARS-CoV-2 and Omicron sequence in different countries for further investigation. Coronanucleoca and DUF5515 were the domains that were targeted for these elements.

#### Multiple sequence alignment and mutation analysis of SARS-CoV-2 and Omicron

Genomic alignments of the SARS-CoV-2 full genome sequence across nations were done using ClustalW [20]. The tool ClustalW was used to align the sequences, which used basic data nucleotides, 0.125 gap extension penalties, and 1.53 opening penalty baseline values. To reduce the number of errors, arrangements with more than 60 problematic bases were eliminated from the data file. SARS-CoV-2 genome was utilized as a reference, and it can be found in GenBank under (AC# NC 045512.2). When the sequences of each wave afterward matched separately, the same alignment procedure was applied. After alignment, the aligned sequences of all genomes were visually reviewed and modified manually by using BioEdit software where small insertions and deletions were eliminated, and

single deletions were filled using the simple majority method. Frame shifts were used to bring the scenes together in the same frame. All sequences, including those with stop codons or frameshift mutations, were aligned.

Multiple sequence alignment and mutation analysis of Omicron was done using the whole sequence of SARS-CoV-2 isolate Wuhan-Hu-1 (NC 045512.2), as the reference genome. Mutations in the full-length genome sequence of Omicron, as well as other variants, were aligned and compared to it using Clustal W.

#### Phylogenetic tree construction of SARS-CoV-2 genome

The phylogenetic relation between Covid 19 sequences from different countries was utilized to generate a phylogenetic tree. In MEGA 7 software, the Neighbour-Joining (NJ) technique was utilized to infer evolutionary history [21]. The proportion of duplicate trees was calculated using 500 bootstrap replicates. Various distance models were employed to determine and assess the genetic distance between nucleotide sequences. The maximum likelihood (ML) trees were produced using the SnpEff method with 500 bootstrap repetitions.

#### Bruvo's distance for spanning network

Minimum spanning networks (MSN) is a great way to visualize relationships among individuals in collected data [22]. Using Bruvo's distance role-based framework, we created a minimal spanning network of selected populations. The minimum spanning network was produced in the software SimuPOP v.1.0.8 [23]. The least spanning network in the data set, with nodes that match MLGs. The demographic participation is represented by the node colors. The width and color of the edges represent distance. A matrix providing the Unicode equivalents of the colors was used in the vertex's colors,

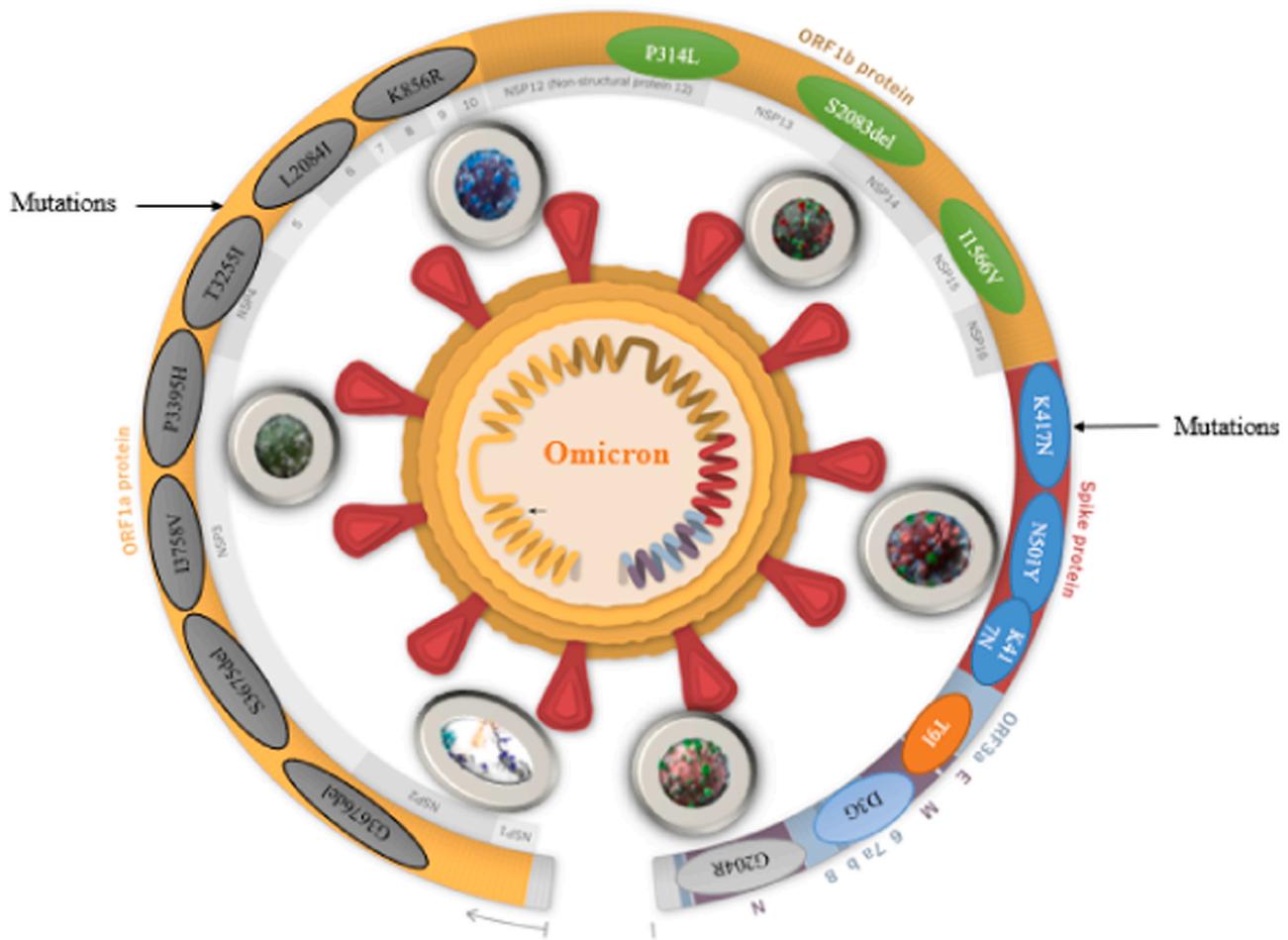


Fig. 1. Genomic structure of Omicron showing the mutational position of different genes and proteins.

along with the vector containing the demographic IDs for the vertex colors. The size of the nodes represents the number of observations. Because fully interactive adjust vertices by diameter, the network node dimensions are expressed by dividing the total number of samples.

#### Genetic diversity analysis

To estimate population genetic statistics, several tools built for the R statistical platform v.3.3.3 were employed [24]. Popper package [25], was used to quantify missing data and test for linkage disequilibrium among loci. Pegas package [26], was used to test for deviations from Hardy-Weinberg Equilibrium (HWE) for each locus and each combination of locus and sampling site using an exact test based on 10,000 Monte Carlo permutations of alleles. P-values were assessed after Bonferroni correction in which the alpha level (0.05) was divided by the number of tests. Popgen report package [27], was used to estimate the frequency of null alleles for each locus [28], private alleles per site, and mean allelic richness using the facton method to correct variation in sample size [29]. Expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), and inbreeding coefficients (FIS) were calculated for each site, and 95% confidence intervals for FIS estimates were determined using 10,000 bootstrap iterations, utilizing the diveRsity package [30], for the calculation. Ade4 package was used to perform a Mantel test with 10,000 permutations to test for genetic isolation by distance. Nei's standard [31], was used to measure the genetic distance to create the genetic matrix.

## Results

### Complete genome information of SARS-CoV-2

The SARS-genetic CoV-2 is composed primarily of 13–15 open reading frames (ORFs) with a total length of 30,000 nucleotides [32]. There are 38% GC and 11 protein-coding regions in the genetic code, with 12 differently expressed genes. SARS-CoV, MERS-CoV, and Omicron contain genomic variation that is nearly identical to ORFs [33,34]. Fig. 1 shows the genomic structure of Omicron. The image also shows the mutational position of different genes and proteins.

### Nucleotide sequence information of SARS-CoV-2 and Omicron among selected countries

The genomic information of SARS-CoV-2 from different countries was retrieved utilizing the GRI database and virus NCBI. Incidences of SARS-CoV-2 and Omicron depending on their complete genomic information are shown in Fig. 2. Around 145 SARS-CoV-2 whole genomes and 12 Omicron full genomes were found from all around the world. These sequences were then analyzed using data mining techniques that were manually applied to conserve domain databases within the NCBI platform to find the most common conserve domains. After data mining, pre-processing, and screening, the most conserved domains in the SARS-CoV-2 and Omicron genome sequences were Corona nucleoca and DUF5515. These two domains were found in every single sequence, which was surprising.

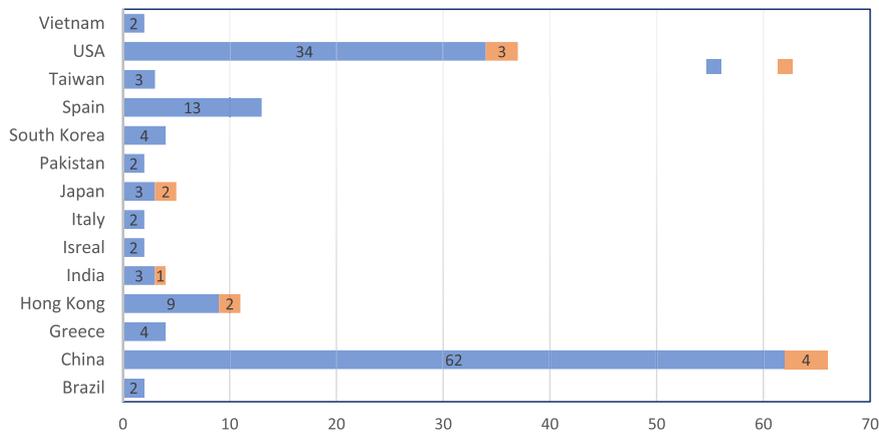


Fig. 2. Details of incidences of SARS-CoV-2 and Omicron depending on their complete genomic information retrieved from the Virus-NCBI database.

Multiple sequence alignment and mutation summary of SARS-CoV-2

To acquire the similarities and dissimilarities in SARS-CoV-2 genome sequences from different countries, we aligned the nucleotide sequence of highly conserved domains viz. Coronanucleoc and DUF5515 with ClustalW and get a plethora of mutations associated with COVID-19 RNA virus from different countries. One mutation was discovered in position 924 (G replaced with A) of the SARS-CoV-2 genome sequence isolated from the Turkish population (Accession No: MT327745.1), while in the United States, T was replaced with C at a location 340 (Accession No: MT345817.1), C was replaced with T at location 571 (Accession No: MT325581.1, MT345817.1, MT304484.1, and MT325596.1, MT325627.1.), C was replaced with T at location 614 (Accession No: MT188339.1), G was replaced with A at location 967 and 1075 (Accession No: MT334529.1 and MN994467.1 and MT325598.1). Although in Pakistan, we observed C/C/G replaced with T/T/A at locations 391, 875, and 924 (Accession No: MT240479\_1). Likewise, in Iran and Hong Kong same mutation was found that G was replaced with A at location 924 (Accession No: MT320891.2, MT114415.1, and MT114414.1). We discovered several mutations in China at various locations. We believe T was changed with C at position 223 (China, Fuyang) under accession no: MT281577.1, although under accession no: MT 123293.2, G was replaced with A at location 155 (China, Guangzhou3). However, T was replaced with C at location 66 in China Hubei, Wuhan6 (Accession No: MT259226.1). Unfortunately, many mutations were discovered in the SARS-CoV-2 genome sequence associated with Wuhan, China, and the expected results are listed in Table 1. Surprisingly, no mutations in the SARS-CoV-2 genome sequence have been found in any other nations. Table 2 shows the detail of mutations in the coding and non-coding regions. Table 3 presents detailed information on the mutations in the putative TRS region (CUAAAC or ACGAAC) of the SARS-CoV-2 strain associated with China. Table 4 shows the annotation data of the SARS-CoV-2, Delta, and Omicron variants from the Wuhan strain with a probable transcription regulatory site (CUAAAC or ACGAAC).

Multiple sequence alignment and mutation summary of Omicron

Previous work found RBD T470-T478 loop and Y505 as viral features for specific detection of SARS-CoV-2 RBD by ACE2. T478 is a common polymorphism in Delta and Omicron variants. RBD has the potential to be developed into an efficient and safe subunit vaccine for SARS-CoV-2 due to its capacity to create exceptionally robust neutralizing antibodies (Nab). Fig. 3 shows a 3D representation of Spike protein in Omicron with probable mutation sites. Omicron has a higher alpha-helix morphology (23.46%) than Delta (22.03%), but less prolonged strands and random coils. The Omicron form of RBD

Table 1

Mutations in the SARS-CoV-2 genome across the globe that may be contributing to the increase in the pathological properties.

Normal nucleotides:

G/C/G/A/C/T/T/T/A/G/C/G/A/T/T/T/A/G/A/A/C/T/G/G/A/C/A/T/G/C/C/T/G/G/A/G/T/C/G/A/G/G/G/A/C/C/T/C/C/T/G/G/A/C/A/A/C/C/G/T/G/G/C/C/G/G/G/C/T/C/C/C/C/T/A/G/G/A/T/A/G/C/T/C/G/G/T/G/T/A/T/C/T/C/A/C/T/G/T/C/C/A/C/A/A/C/G/A/T/A/T/A/G/G/C/C/T/G/A/A/C/A/G/A/C/T/G/A/G/T/G/C/G/G/A/C/G/C/G/A/A/A/G/C

Mutated nucleotides:

T/T/C/C/A/A/C/G/C/C/C/A/C/A/C/C/T/C/G/T/G/T/T/G/T/T/C/A/A/C/A/A/G/T/T/A/T/T/T/G/A/A/C/G/A/C/G/A/C/G/A/A/T/A/C/T/G/A/A/T/C/C/R/G/T/G/G/G/T/A/A/A/A/T/G/A/T/A/T/A/G/T/A/A/G/T/A/C/C/A/T/G/T/T/C/C/A/A/G/A/T/A/A/C/A/G/A/C/G/T/C/T/G/G/A/T/G/C/G/A/T/G/A/A/G/A/A/G/C/T/G/T/G/T/C/T/A/C/T/G/C/T/G/G/T/T/C/T/C/G/T/G

Genomic location of residues to be mutated:

215,216,221,222,224,226,227,229,231,233,234,236,237,238,241,242,243,244,-246,247,248,349,250,251,260,261,262,264,265,267,270,271,273,275,276,277,-278,282,283,284,285,289,291,294,298,300,304,306,307,308,309,311,313,316,-318,319,320,324,325,328,331,334,336,337,338,342,344,345,346,347,348,349,-352,355,356,357,358,360,361,362,364,365,368,369,371,373,374,378,379,382,-383,385,388,392,393,398,399,400,406,407,409,411,412,413,415,417,418,419,420,422,424,425,426,427,429,435,436,437,440,459,462,465,466,467,469,470,-471,472,473,474,476,477,479,481,482,485,488,490,492,498,499,500,502,503,-504,507,509,510,512,515,516,517 (Accession No: MT291828.1).

has a greater alpha helix concentration (8.30%) than the Delta variant (5.68%). Although the expected increase in alpha-helices indicates that alpha helices are more resistant to mutations than beta strands, Omicron has a far larger conformational change content.

According to multiple alignments, we identified a Spike protein mutation in Delta and Omicron variants that are differentially

Table 2

Detail of the mutations in CDS and Non-CDS regions of SARS CoV-2.

S. No	Gene name	Protein Name	GB ID	Number of Mutations
1	ORF1ab	orf1ab polyprotein	QHD43415.1	321
2	S	Spike surface glycoprotein	QHD43416.1	51
3	ORF3a	ORF3a protein	QHD43417.1	36
4	E	Envelope protein	QHD43418.1	2
5	M	membrane glycoprotein	QHD43419.1	5
6	ORF6	ORF6 protein	QHD43420.1	1
7	ORF7a	ORF7a protein	QHD43421.1	0
8	ORF8	ORF8 protein	QHD43422.1	50
9	N	Nucleocapsid phosphoprotein	QHD43423.2	60
10	ORF10	ORF10 protein	QHI42199.1	2
11	Non-CDS	Non-CDS	Non-CDS	68
<b>Total</b>				<b>596</b>

**Table 3**  
Details of the mutations in the putative TRS region (CUAAAC or ACGAAC).

S.No	Strain	Location	Original residue	Replaced residue	TRS of Gene
1	ChinaHubeiWuhan6 (MT259226.1)	27389	T	C	ORF7a protein
2	ChinaGuangzhou1 (MT123291.2)	29532	G	A	ORF10 protein
3	ChinaGuangzhou3 (MT123293.2)	29532	G	A	ORF10 protein

expressed. The Omicron variant has 30 mutations in the Spike protein, half of which are in the RBD (Table 5). Delta variant, which has a sequential NCBI ID QWK65230.1 has genetic changes D614G, P681R, D950N, T19R, G142D, 156–157, R158G, 213, and Omicron variant, sequential ID from NC N440K, G446S, S477N, T478K, E484A, Q493R,

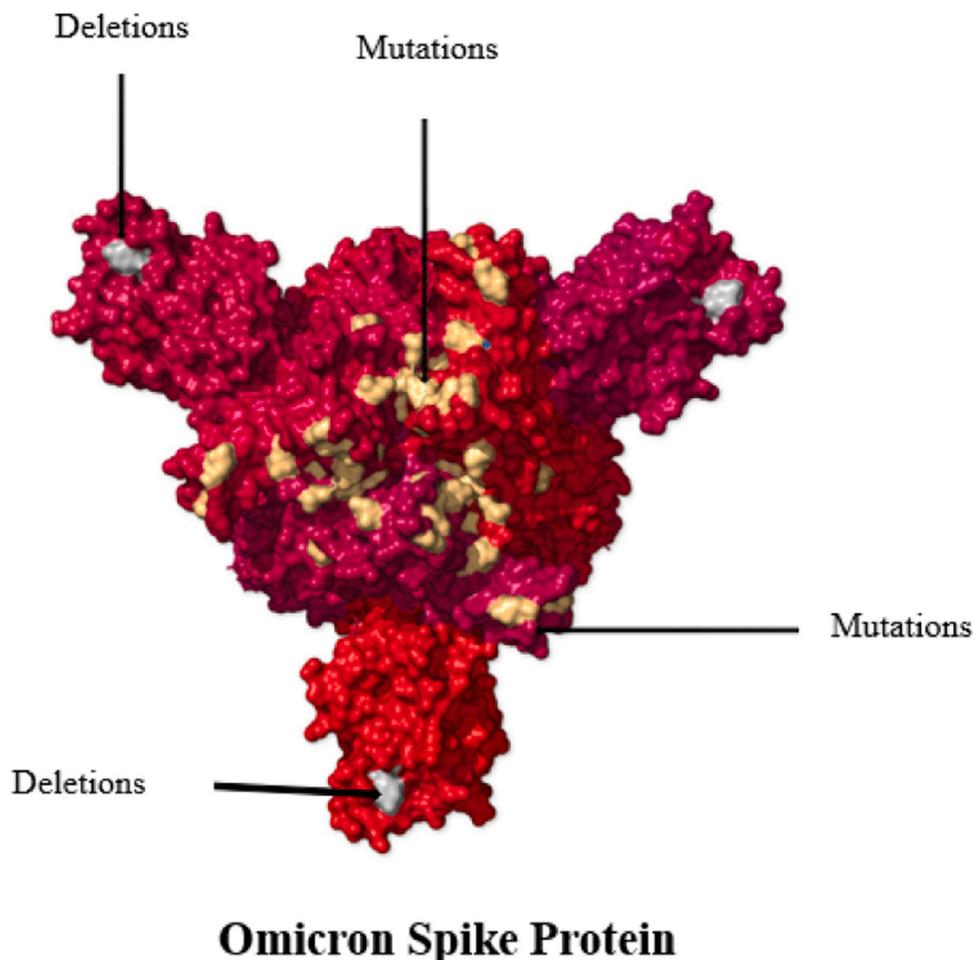
G496S, Q498R, N501Y, Y505H, G339D, S371L, S373P, S375F, K417N are mutations in Omicron. Mutations in Omicron's spike protein K417N, D614G, N501Y, K417N, G142D, T478K, N501Y, and D614G have all been found in other viral kinds of concern and are thought to make the virus more infectious. Fig. 4 shows mutational rates in Omicron and other variants. The average number of mutations is highest in Omicron (54%) followed by Delta (36%), Alpha (30%), Gamma (29.5%), and Beta (29%) mutations.

*Phylogenetic analysis of SARS-CoV-2 genome among different countries*

The phylogenetic tree of Covid 19 sequences among different countries is shown in Fig. 5. The red open circle represents the subclade's four constituents. The blue open circle represents two elements in Subclade 5. A lime open circle represents four elements

**Table 4**  
Annotation details of SARS CoV-2, Delta and Omicron variants from wild type Wuhan strain, with putative transcription regulation site (CUAAAC or ACGAAC).

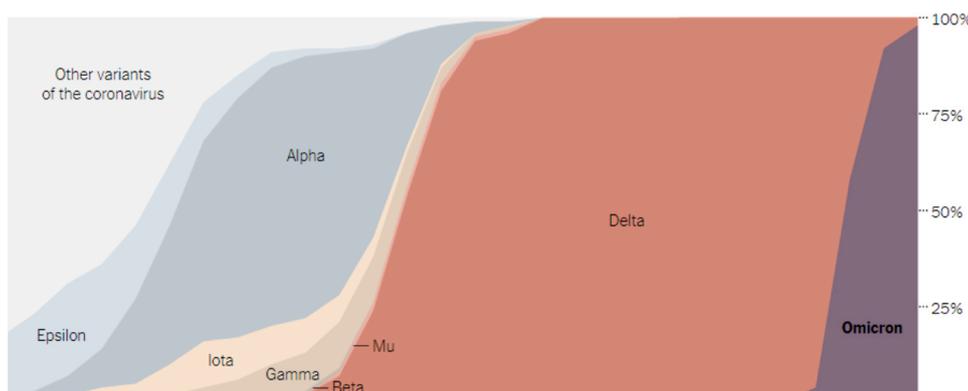
S. No	Gene	Location nt (start and end)	Protein	TRS Location with putative TRS size	Mutations in TRS site
1	ORF1ab	266–21555	orf1ab polyprotein	66 (10)	0
2	S	21563–25384	Spike surface glycoprotein	21552 (10)	0
3	ORF3a	25393–26220	ORF3a protein	25385 (6)	0
4	E	26245–26472	Envelope protein	26237 (10)	0
5	M	26523–27191	membrane glycoprotein	26469 (6)	0
6	ORF6	27202–27387	ORF6 protein	27041 (6)	0
7	ORF7a	27394–27759	ORF7a protein	27388 (6)	1
8	ORF8	27894–28259	ORF8 protein	27884 (10)	0
9	N	28274–29533	Nucleocapsid phosphoprotein	28256 (10)	0
10	ORF10	29558–29674	ORF10 protein	29530 (6)	2



**Fig. 3.** 3D representation of Spike protein in Omicron with probable mutation sites.

**Table 5**  
Mutations in Omicron genome variant across the globe that may be contributing to the increase in the pathological properties.

Gene/protein	Mutations	Conserved mutations
S	G142del, V143del, Y144del, Y145D, N211del, L212I, ins214E, ins215P, ins216E, G339D, S371L S373P, S375F, K417N	K417N, D614G, N501Y, K417N.
S1/ RBD	A67V, T95I, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, E484A, Q493K, G496S, Q498R, Y505H, T547K, N501Y	G142D, T478K, N501Y, D614G,
S2	H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F, N764K, D796Y, N856K, Q954H, N969K, and L981F	G142D, T478K, N501Y, D614G,
RBM	K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505, N440K, G446S, Q493R, G496S, Q498R, and Y505H	K417N, N440K, G446S, Q493R, G496S, Q498R, and Y505H
ACE2	G446S, Q493R, G496S, Q498R, and Y505H,	
ORF1a	K856R, L2084I, A2710T, T3255I, P3395H, L3674del, S3675del, G3676del, and I3758V	
ORF1b	P314L and I1566V	
NSP9b	P10S, E27del, N28del, and A29del	
E	T9I	
M	D3G, Q19E, and A63T	
N	P13L, E31del, R32del, S33del, R203K, and G204R	



**Fig. 4.** Graph showing rates of mutations in Omicron and other variants of SARS-CoV-2.

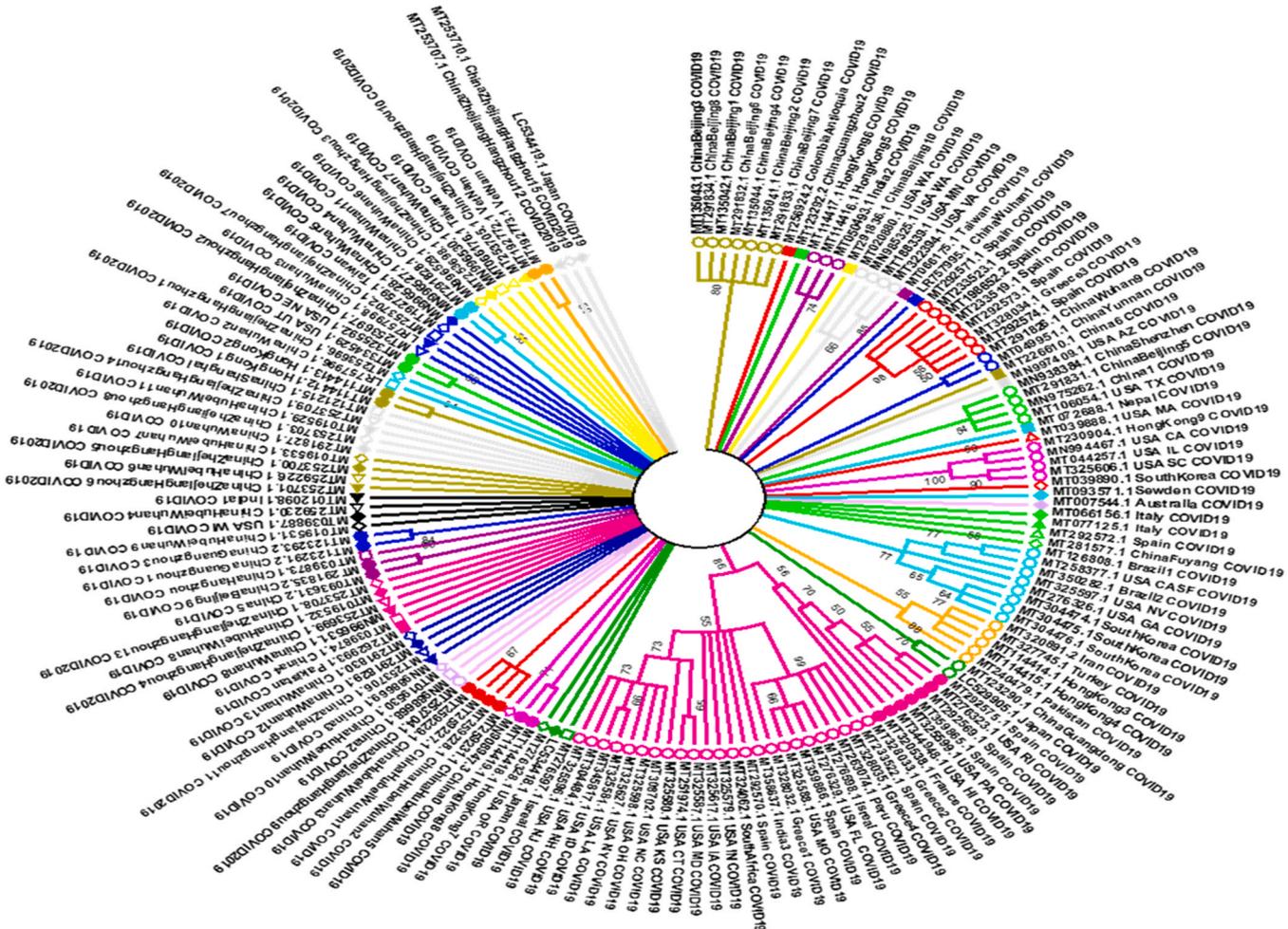
in Subclade 6. The ultra-open circle represents four elements in Subclade 7. A cyan open circle represents the eight elements of Subclade 8. The orange open circle represents Subclade 9, which has five elements. The green open circle represents two elements in Subclade 10. A pink open/filled circle represents Subclade 11's 31 elements. Two elements of Subclade 12 are represented by an ultra-filled circle. A red-filled circle represents the four elements of Subclade 13. A purple-filled circle represents two elements in Subclade 14. A blue-filled circle represents two elements in Subclade 15. Brown-filled circles represent two elements in Subclade 16. A lime-filled circle represents two elements in Subclade 17. The blue circle represents two elements in Subclade 18. Two elements of Subclade 19 are represented by an orange-filled circle. The remaining elements are represented by different colors, colors, and forms to identify them from one another.

A sub-clade distribution was visible on the tree (Fig. 5). For a total of 157 sequences, each sub-clade represents the grouping of sister families. The members of the China Beijing family were grouped in one sub-clade. As a result, HongKong6, HongKong5, and India2 belong to sister families, whereas China Beijing10, USA-WA1, USA-WA2, USA-MN, USA-VA, and Taiwan belong to the same family. The Spain family was grouped into a single sub-clade. China6 and USA-AZ created a sister family. China Shenzhen, Beijing5, China1, and USA TX were all in the same cluster. Nepal and USA MA have the same family. The largest group consists of 38 components divided into six sub-clades; the first sub-clade consists of five elements shared by Iran, Turkey, Hong Kong3, Hong Kong4, and Pakistan. The second sub-clade has two elements that belong to the same family as China Guangdong (MT123290.1) and Japan (LC529905.1). France, USA RI, USA PA, USA HI, and Spain accession "MT292575.1, MT359865.1, and MT292569.1" are clustered together with the same group in the

third sub-clade. The fourth sub-clade contains nine elements, where Greece1, Greece2, Greece4, Israel, Peru, USA\_MO, USA\_FL, and Spain accession MT233522.1 and MT359866.1 clustered in the same group representing their respective families. The fifth sub-clade contains five elements, where India3, Spain, South Africa, USA\_IN, and USA\_IA, share the same group. Conclusively, the last sixth sub-clade contains ten elements: USA\_MD, USA\_CT, USA\_KS, USA\_NC, USA\_NY, USA\_OH, USA\_LA, USA\_ID, USA\_NH, and USA\_NU clustered together in the same group. VitNam1 and VitNam2 constitute sister families with Taiwan and share the same group. ChinaHubeiWuhan5 (MT259231.1), ChinaHubeiWuhan2 (MT259228.1), ChinaHubeiWuhan1 (MT259227.1), ChinaHubeiWuhan3 (MT259229.1), and ChinaZhejiangHangzhou15 clustered in one group. ChinaWuhan10, ChinaGuangzhou3, ChinaGuangzhou1, ChinaHubeiWuhan9, ChinaHubeiWuhan4, USA\_WI, and India grouped.

*The phylogenetic relations of Omicron and other variants*

The research studies of Omicron mutations were identified using a variety of adaptive mutation methodologies. According to vulnerable & criterion cluster analysis strategies, the Omicron modified version forms a new polyphyletic clade that is distanced from other SARS-CoV-2 versions. Using a basic nucleotide substitution model like Jukes-Cantor, the NJ approach discovered a close relationship between Omicron variations and the newly discovered Alpha variant. The percentages of sequence homology among the Omicron, Alpha, Gamma, Delta, Beta, and SARS-CoV-2 USA isolates all were comparable as shown in Fig. 6. According to a genome comparison with other variations, the Omicron variant's genome has the greatest gaps, ranging from 45 to 65. Even though they founded their group, Omicron is likely to have lived considerably longer than normal due



**Fig. 5.** Phylogenetic tree of Covid 19 sequences from different countries. The tree was constructed by the Neighbor-Joining method with 500 bootstrap replicates using the MEGA7 program. The bootstrap support values (%) are shown near the nodes, and values > 50% are not shown. Representing 157 total sequences, each sub-clade represents the clustering of sister families.

to their proximity to the Alpha variant. Detailed and continuous sequence investigations are critically focused on understanding the virus's evolution and mutations.

*Bravo's distance and Population outcomes*

The findings are anticipated by the minimal spanning tree, which maximizes the lengths (or "weights") of the vertices of the sampling population tree. A rule-based strategy produced a minimal spanning matrix of the chosen population, as shown in Fig. 7. Each component represents a different multi-locus genotype. Edge thickness and coloration show interconnectedness, while node colors represent demographic engagement. The thickness and hue of the edges are proportionate to 'Bruvo's distance, but the length of the edges is random.

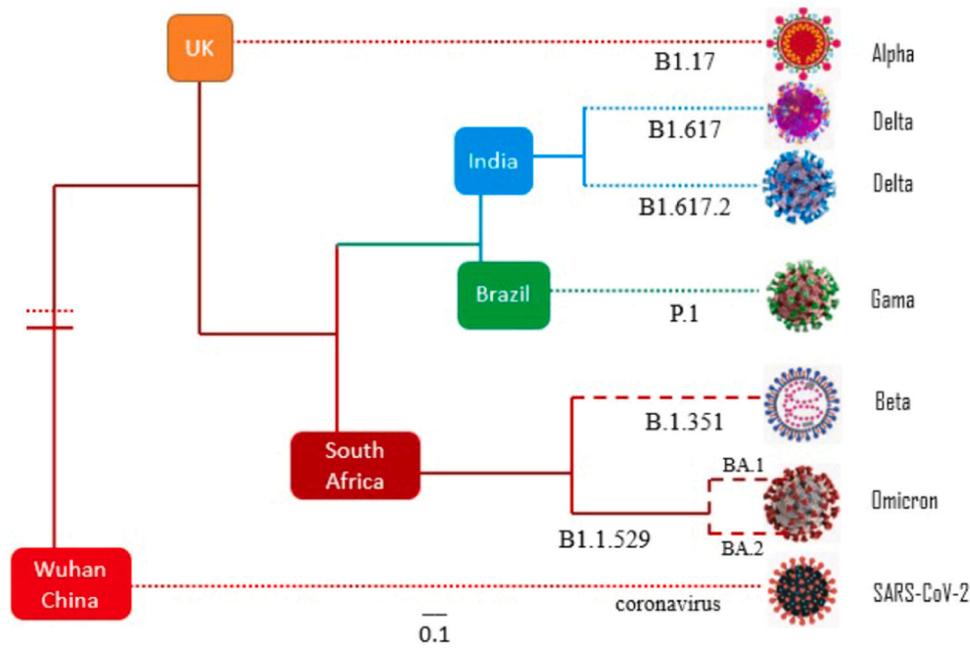
*Population genetics analysis*

The Euclidean coefficient was used to examine the genetic connection of SARS-CoV-2 among 157 strains from 20 countries. Based on the sequence of the whole viral genome among 157 genomes, the dendrogram was analyzed and created. Based on the similarity matrix, the UPGMA dendrogram demonstrates a clear separation between the SARS-CoV-2 genome sequences of different nations. Different clusters were created using cluster analysis [35]. In Fig. 8, an interconnected genomic sequence is put in a similar cluster.

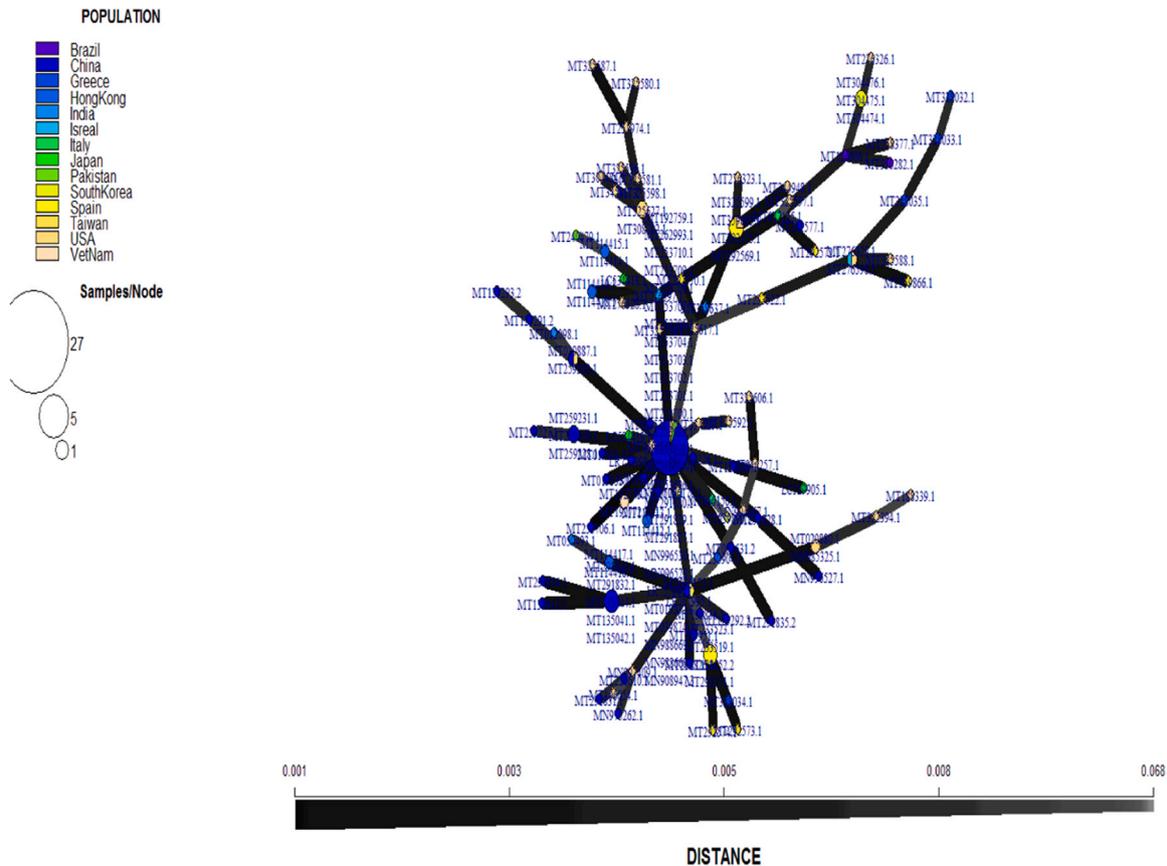
Abundant genetic variability can strengthen a 'species' ability to respond to changing environments and consequently enhance its evolutionary potential. Genetic diversity is affected by many factors such as breeding system, seed dispersal, life form, geographical distribution, and historical origin [36,37]. Nei's Genetic Distance is a measure of genetic divergence across species or populations within a species, whether defined in time from a common ancestor or in the degree of differentiation. As demonstrated in Fig. 9, the concept of genetic distance is also used to understand the origins of biodiversity.

Analysis of molecular variance (AMOVA) was used to find variance within-sample, between the sample and the population [38]. It was performed in the R package by splitting the genotypes across all loci to create multiple haplotypes, which helped calculate within-sample, within the population, and between-sample distances and incorporate them into the model. For all genomic sequences, the mean  $H_o$  (Observed Heterozygosity) was 0.01, and the mean  $H_e$  (Expected heterozygosity) was 0.32, respectively. The overall  $H_t$  (total genetic diversity index) was 0.01943989 while  $H_s$  (gene diversity within populations) was 0.01266951.  $G_{st\_est}$  (Coefficient of genetic differentiation) was 0.52324978,  $D_{het}$  (average heterozygosity) was 0.01371452, and  $G_{prime\_st}$  (Hedrick) was 0.52324978.

The variation with samples was relatively low compared to variation between samples and population. For variation within samples, the observed gene diversity within samples was 0.07586207, while the standard gene diversity index was -39.8514, leading to a significant loss of genetic diversity within samples ( $p=0.01$ ). For



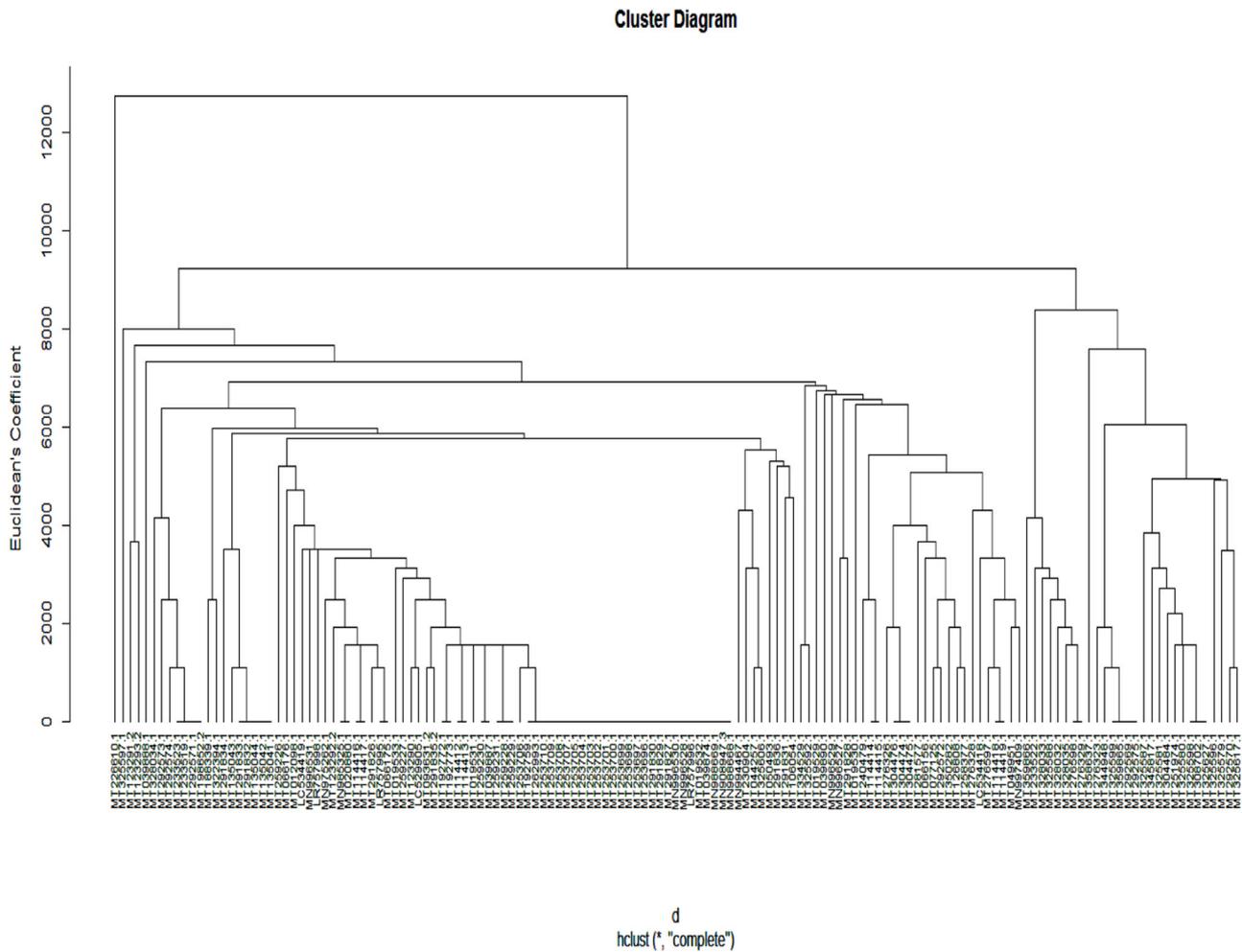
**Fig. 6.** Phylogenetic relation of Omicron and other variants. A basic Jukes-Cantor and NJ approach discovered percent sequence homology between Omicron and other SARS-CoV-2 variants.



**Fig. 7.** Spanning network representing 157 genomes of different countries using 'Bruvo's distance. Each node represents a unique multilocus genotype. Node shading (colors) represent population membership, while edge widths and shading represent relatedness. Edge length is arbitrary.

variation between samples, the observed genetic diversity was 6.16872760, and the standard gene diversity index was 23.60827, resulting in significant genetic differentiation among various sequences ( $p = 0.01$ ). For variation between populations, the observed

genetic diversity was 1.31912384, and the standard gene diversity index was 10.77174, resulting in significant genetic differentiation among various populations ( $p = 0.01$ ) are mentioned in Table 6 and Fig. 10.

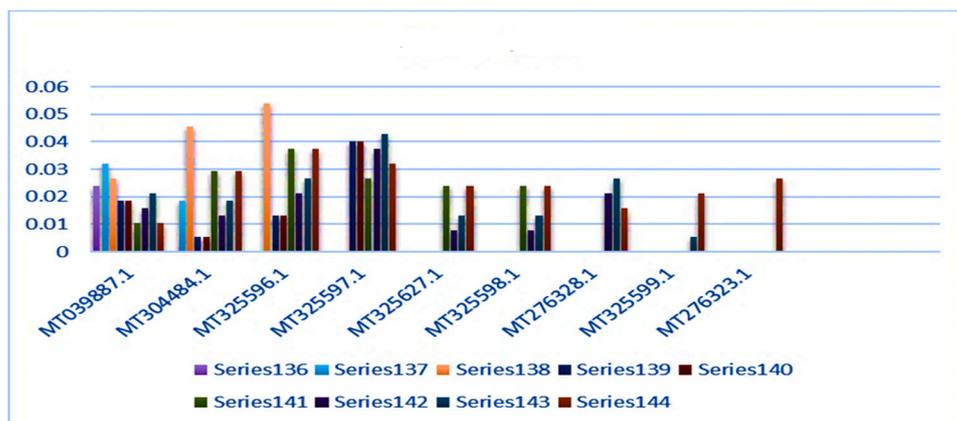


**Fig. 8.** UPGMA dendrogram based on the Euclidean coefficient representing the genetic similarity between the genome of SARS-Co-2 among different countries. Bootstrap values greater than 25% are mentioned.

**Discussion**

The COVID-19 epidemic has hit each region, generating a major public health crisis that has been exacerbated by the invasion of many viral strains. According to the WHO, the Omicron variant of SARS-CoV-2 has been detected in at least 63 countries [39]. With the rapid development of SARS-CoV-2 genomic sequence data and its variant Omicron, it is critical to successfully predict mutations

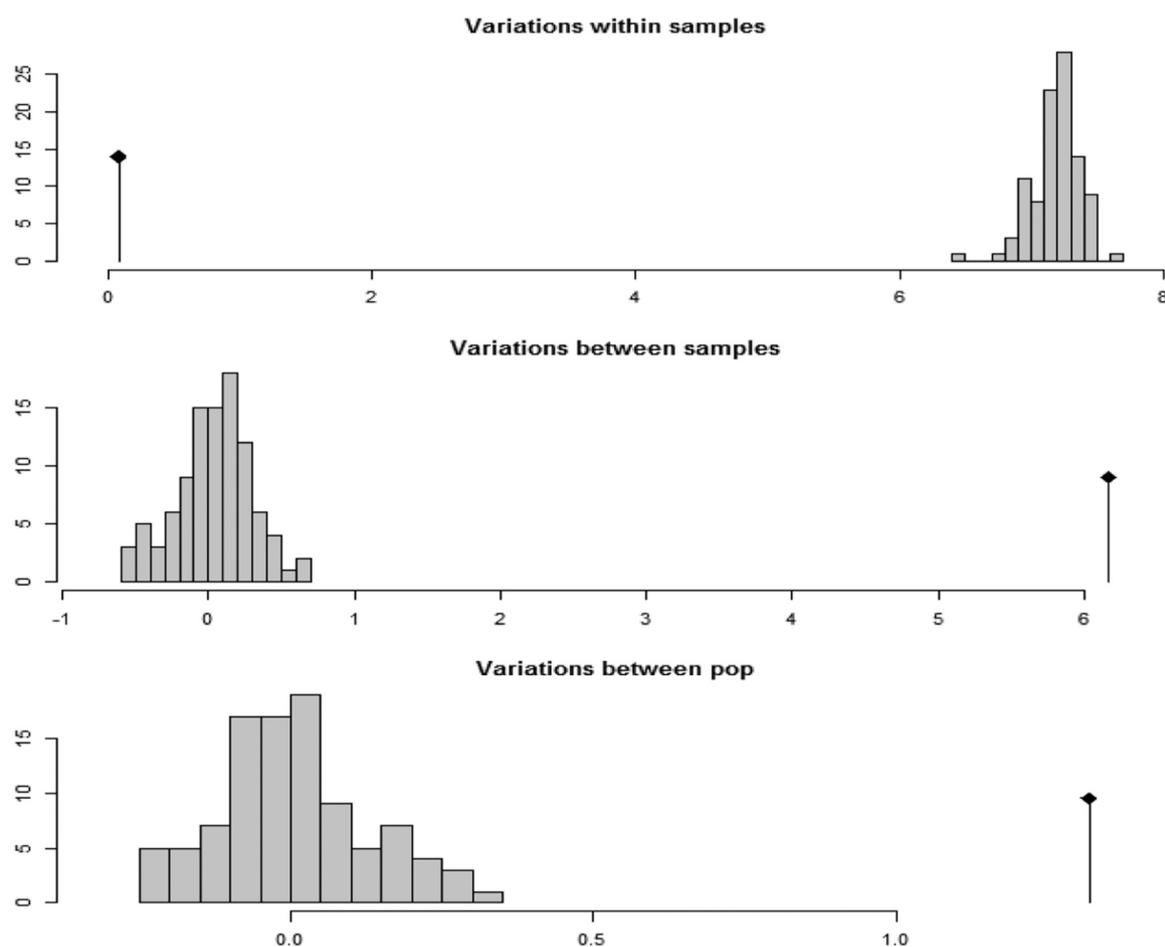
within the genome. Several investigations are being conducted on spike protein to develop vaccines and drugs against SARS-CoV-2 or Omicron, as it is responsible for viral interaction with the host cell membrane and viral entry into the cell. In this study, we used a range of methods to gain an overview of the development of the SARS-CoV-2 and Omicron in the context of mutation screening and phylogenetic and genetic diversity analysis among different countries. The present study comprised SARS-COV-2 and Omicron sequences



**Fig. 9.** Genetic diversity based on Nei's coefficients among highly mutated sequences.

**Table 6**  
Detailed information of all selected populations of SARS-CoV-2 in the context of a range of techniques applied by statistical analysis related to population genetics.

Populations	Number of individuals /Isolates (N)	Number of Multilocus Genotypes (MLG)	Expected number of MLG (eMLG)	Standard Error (SE)	Shannon's Index (H)	Stoddart and Taylor's Index (G)	Simpson Index (lambda)	Evenness (E.5)	Nei's Gene Diversity (Hexp)	Index of Association (Ia)	Standardized Index of Association (rbarD)
Brazil	2	2	2.00	0.000	0.693	2.00	0.500	1.000	0.0035	NA	NA
China	62	32	6.69	1.398	2.646	5.59	0.821	0.350	0.0110	6.24e+ 00	7.99e-02
Greece	4	4	4.00	0.000	1.386	4.00	0.750	1.000	0.0225	9.80e+ 00	5.16e-01
Hong Kong	9	5	5.00	0.000	1.581	4.76	0.790	0.975	0.0114	1.39e+ 00	1.16e-01
India	3	3	3.00	0.000	1.099	3.00	0.667	1.000	0.0266	-6.32e-01	-3.51e-02
Israel	2	2	2.00	0.000	0.693	2.00	0.500	1.000	0.0157	NA	NA
Italy	2	2	2.00	0.000	0.693	2.00	0.500	1.000	0.0035	NA	NA
Japan	3	3	3.00	0.000	1.099	3.00	0.667	1.000	0.01	12 - 8.75e-01	-1.25e-01
Pakistan	2	2	2.00	0.000	0.693	2.00	0.500	1.000	0.0105	NA	NA
South Korea	4	2	2.00	0.000	0.562	1.60	0.375	0.795	0.0135	1.10e+ 01	1.00e+ 00
Spain	13	8	6. 61	0.784	1.885	5.45	0.817	0.797	0.0215	5.54e+ 00	2.46e-01
Taiwan	3	3	3.00	0.000	1.099	3.00	0.667	1.000	0.0056	4.44e-16	1.25e-16
USA	34	31	9.70	0.525	3.389	27.52	0.964	0.927	0.0246	1.48e+ 00	2.10e-02
Vietnam	2	1	1.00	0.000	0.000	1.00	0.000	NaN	0.0000	NA	NA
Total	145	94	8.85	1.074	4.077	22.68	0.956	0.374	0.0189	2.59e+00	1.55e-02



**Fig. 10.** Analysis of molecular variations with samples, between samples, and with population.

from 157 strains as well as whole-genome sequencing from various nations, including China, Germany, the United States, Pakistan, and Turkey. We gathered sequences from Brazil, Australia, Vietnam, Pakistan, Peru, Nepal, Japan, the United States, Taiwan, Sweden, Italy, and South Korea based on the findings of the research. This mutation database enables users to easily find and analyze mutations in the COVID-19 and its variants omicron genome by conducting alignment and constructing a phylogenetic tree. It may be used to analyze variation between closely related sequences and variety as

molecular markers, as well as to annotate and categorize numerous comparable sequences.

The genetic evidence of any life is stored in its genome and annotation is the first step in deciphering the sequence. The SARS-CoV genome is approximately 30 kilobytes in length. Only a few coding genes, however, do not appear to agree with the viral genome's general assets and the least grouping of hereditary data. It also contains several non-structural proteins; yet, having all of the information in one location is crucial. The absence is most likely the

result of their short prevalent time before disintegration. The genetic variance between the freshly sequenced genomes of SARS-CoV-2 was defined by an earlier study in 2020 [40]. The genomic sequence analysis of pathogenic organisms is becoming more widely recognized as a critical technique for understanding infectious diseases [41]. Whole-genome sequence evidence illuminates critical epidemiological constraints including outbreak/epidemic survival time, transmission path reconstruction, and evaluation of possible causes and animal reservoirs.

The Omicron variant's first sequenced genomes were utilized to trace its evolutionary relationships with other SARS-CoV-2 variants. The sample was taken in Botswana on November 11, 2021, and the nucleotide sequence was submitted with the AC# EPI\_ISL\_6640916 on November 23, 2021. Alpha, Beta, Gamma, Delta, Mu, GH49R, and the SARS-CoV-2 USA isolates were among the other variants. The variants' FASTA files were aligned with Clustal Omega, and the results were compared for similarity confidence [42]. The Omicron variation has the largest genetic diversity, according to the reciprocal matrix, as we discovered in the "population genetics analysis" section in Results. The Omicron genome has the following number of nucleotide changes when compared to other variants: SARS-CoV-2 USA isolate > Mu variant > Beta variant > Delta variant > Gamma variant > Alpha variant > Omicron variant including 141, 140, 138, 132, 130, and 109 mutations, which are identical to our results [43]. The Alpha variant has the highest identification probability (99.63%) with the Omicron version, trailed by Gamma & Mu variants (99.56%). The SARS-CoV-2 isolated from the United States has the lowest identity (99.53%) as per our results [43]. Similarly, during genomic identification with several other viruses, the Omicron variety had the most gaps, ranging from 45 to 65, all ranging conformation matched with earlier research [43].

The available genomic data were used to create phylogenetic trees based on the distance matrix, where Wuhan China is the origin of the coronavirus and the closest distance of 0.1 was used to predict relationships as per earlier research [44], where we found SARS-CoV-2 is the origin of other variants mapped in different countries. According to a phylogenetic tree generated using vulnerable relationships between varieties and similar evolutionary speeds throughout branches using the UPGMA algorithm as per the methodology of research used in phylogenetic tree investigation [45], the Omicron variety is evolutionarily distant from other variants, resulting in a monophyletic clade [46]. In the present study, the UPGMA method combines pairs of nucleotides separated by a small distance, and it was evident that the Omicron variation was distinct from the others. The earlier study also looked into the Omicron variant, which was found to be distinct from other variations, and created a separate polyphyletic classification based on Kimura 80's NJ methodology, which successfully addresses genetic variants as well nucleotide transitions and transversions [43]. Using a simplified model of JC replacement with the same probable nucleotide mutation probability, a close association with the Alpha modified version was discovered.

The Omicron variant reveals an unusually high prevalence of infection and re-infection [47]. The Wuhan-Hu-1 variety, which was classified as an Alpha lineage, increased disease transmission by 40–90%, while the Delta lineage was even more infectious, and the Omicron VoC was the most virulent [48]. Furthermore, the Omicron lineage has a 5.4-fold higher re-infection rate than the Delta lineage, which is the most contagious [49]. According to our Spike study, this improvement in disease transmission is due to more than just the Spike RBD's higher affinity for the human ACE2 receptor. The RBD binding affinity of the Omicron for human ACE2 receptors is 2.4 times more than the Wuhan-Hu-1 strain, according to pseudovirus experiments [50], which is lower than Alpha (6.2 times higher than Wuhan-Hu-1), similar to Beta (2.4 times higher than Wuhan-Hu-1), and higher than Delta (1.2 times higher than Wuhan-Hu-1). Because

it emerged in an environment where several participants had already been exposed to the virus by other varieties or were immunized to a significant degree by a significant rise in immunization, the high infection and re-infection rate of Omicron could be attributed, at least in part, to strong development to deliver immunogenic transition. The Omicron lineage has previously evaded 26 of the 29 monoclonal antibodies aimed toward the Spike RBM [51,52].

Our data of 157 distinct SARS-CoV-2 viral strains including Omicron variants and entire genome sequences from different countries utilizing numerous computational methods showed conserved areas in a protein alignment for homology patterns to identify diverse lineages. Corona nucleoca and DUF5515 were the most conserved domains in every single sequence. In comparison to the Wuhan-Hu-1 strain, all genomes gained mutations, particularly in the TRS region (CUAAAC or ACGAAC) of the SARS-CoV-2 strain connected with China. Each node of the spanning network in 'Bravo's distance represents a separate multi-locus genotype which was closely related to Delta and Omicron variants. The Omicron has a stronger alpha-helix morphology (23.46%) than the Delta variant (22.03%), Omicron variant has 30 mutations in the Spike protein. We also identified that T478 is a common polymorphism in Delta and Omicron variants and genomic gaps, ranging from 45 to 65aa. Large alterations in the RBD area of the Omicron variant may contribute to the high binding specificity with hACE2. A morphological study revealed that all 157 sequences had variations and comply with Nei's Genetic distance. The AMOVA results revealed that heterozygosity ( $H_s$ ) was 0.01, mean anticipated  $H_s$  was 0.32, genetic diversity index (GDI) was 0.01943989, and GD within the population was 0.01266951. The GD coefficient was 0.52324978, the average  $H_s$  was 0.01371452, and the Hedrick value was 0.52324978. Brazil has the largest standard error (SE) rate (1.398) among nations, whereas Japan has the highest ratio of Nei's gene diversity (0.01).

The COVID-19 outbreak has enabled ongoing research engagement and data exchange, allowing scientists to examine SARS-CoV-2 recombination and propagation activities in real-time [53]. Epidemiologic, genomic, and functional studies of variation must be effectively utilized to determine how and where to slow down and quickly eradicate within- and between-species transmissions [54]. Genetic variations, insertions/deletions, homologous intertypic, and possibly non-homologous reintegration (gene duplications, horizontal gene transfers) have all demonstrated extraordinary evolutionary plasticity in a variety of CoVs. Based on these findings, at least five possible evolution scenarios have been presented, each of which could have an impact on the progression of the COVID-19 pandemic or the emergence of another highly contagious CoV in the future. According to our results, SARS-CoV-2 developed through inhabitants' structure, because lineages are formed up of ancestors that are either VOCs or VOIs. Biological classification into phylogenetic figure clades provides a very coarse description that may lack appropriate accuracy for clinical connection. Furthermore, as per the same consequence, when new lineages are found and the clinical impact of extant genealogies on pathogenicity is considered more seriously, lineage diversity may alter over time and makes different variants of SARS-CoV-2 like Alpha, Beta, Delta and Omicron, etc. In this research, we employ different technique and protocols which provides crucial justification of the basic lineages, as well as the geographical spread of the variants as per selected regions like Wuhan China, UK, South Africa, India, and Brazil, etc. The contemporary epidemic, rapid data acquisition, & exponential quantitative investigations present several chances to build gold-standard scientific strategies for the planning and implementation of epidemic procedures to avoid future outbreaks. The development of the Omicron VoC, as well as the diversification of the Coronavirus sub-family as a whole, serves as a warning to researchers, manufacturers of pneumococcal conjugates, and legislators that they must remain cautious. More significantly, given the Spike ORF's naturally high

mutation and crossover rate, as well as the already licensed vaccination schemes and immunotherapies that primarily target this area, other more stable genetic areas should be studied as future targets. Such a multifaceted approach would've been comparable to HIV treatment plans, in which combination medications have largely superseded the use of single drugs, which the virus has immediately replied to.

To assess the influence of mutations on vaccination effectiveness, inhibitory studies were done on true viruses as well as pseudoviruses with particular Omicron mutations, as well as wider sets of mutations suggesting various variations of concern and other circulating genetic alterations [51]. The adaptability of harmful pathogens is enhanced by evading human immune response and defining making them extremely infectious. Contempt becoming a strong likely mutation possibility discovered in the many places included in the study area, additional scientific and practical research is needed to determine the specific importance of this somatic mutation in relation to molecular pathways and interactions in biological systems. The findings of the study potentially benefit the research community in better understanding the genetic evolution of SARS-CoV-2 and the variant Omicron. Our computational-based approaches provide a quick and low-cost method for predicting the biochemical consequences for developing a new vaccine. This anticipation may also offer the thinktank with a fantastic chance to perform a future study.

## Conclusion

In the current study, complete mutation analyses, genomic analysis, population genetic study, genetic diversity approach, and molecular variance are presented, providing a snapshot of rapidly changing genomic mutation in the SARS-CoV-2 genome including Omicron variants among various countries. However, given the crucial importance of identifying potential signatures of adaptation in SARS-CoV-2 for guiding the ongoing development of vaccines and treatments, we propose a plausible approach and initial list to facilitate future work and interpretation of the observed patterns. More data continues to be made available, which will allow an ongoing investigation by us and others. We believe it is vital to continue monitoring the SARS-CoV-2 evolution in this way and make the results available to the scientific community. In this framework, we hope that the interactive research we provide will help identify critical recurrent vaccine designing strategies against the pathogenic nature of SARS-CoV-2.

## Funding

This work was funded by the Researchers Supporting Project number (RSP-2021/339), King S aud University, Riyadh, Saudi Arabia.

## Ethics Approval and Consent to Participate

Not Applicable.

## CRedit authorship contribution statement

All authors contributed to the study's conception and design. A.A.K., I. Z. & B. H. K., supervised this manuscript, and finally conceived and designed the study protocols. Material preparation, data collection, analysis, and manuscript writing were performed by M. A., B. H. K., S. U. A., Z. J., Y. A., A.M.A., A.M., A.A. A. A. K., and M. A. R. reviewed and proofread the manuscript. All authors read and approved the final manuscript.

## Data Availability

Not applicable.

## Declaration of Competing Interest

We have no conflict of interest to declare.

## Acknowledgements

This work was funded by the Researchers Supporting Project number (RSP-2021/339), King Saud University, Riyadh, Saudi Arabia.

## Consent for Publication

Not applicable.

## References

- [1] Li J-Y, et al. The epidemic of 2019–novel-coronavirus (2019-nCoV) pneumonia and insights for emerging infectious diseases in the future. *Microbes Infect* 2020;vol. 22(2):80–5.
- [2] Zhou P, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* 2020;vol. 579(7798):270–3.
- [3] Wu C, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;vol. 180(7):934–43.
- [4] Jee Y. WHO international health regulations emergency committee for the COVID-19 outbreak. *Epidemiol Health* 2020;vol. 42.
- [5] Van Dyk AC. HIV/AIDS Care and Counselling: A Multidisciplinary Approach. South Africa: Pearson; 2010.
- [6] Bhat EA, et al. SARS-CoV-2: insight in genome structure, pathogenesis and viral receptor binding analysis—an updated review. *Int Immunopharmacol* 2021;vol. 95:107493.
- [7] Bermúdez JMG, et al. Mutation profile of SARS-CoV-2 genome in a sample from the first year of the pandemic in Colombia. *Infect. Genet Evol* 2021:105192.
- [8] Virtanen A, Henriksson N, Nilsson P, Nissbeck M. Poly (A)-specific ribonuclease (PARN): an allosterically regulated, processive and mRNA cap-interacting deadenylase. *Crit Rev Biochem Mol Biol* 2013;vol. 48(2):192–209.
- [9] Meyer KD, Jaffrey SR. The dynamic epitranscriptome: N 6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol* 2014;vol. 15(5):313–26.
- [10] Tripathi S, et al. Novel coronavirus (Sars-cov-2): molecular biology, pathogenesis, pathobiology and advances in treatment of covid-19 patients—an update. *J Exp Biol Agric Sci* 2020;vol. 8(6).
- [11] Rottier PJ. The Coronavirus Membrane Glycoprotein. in *The Coronaviridae*. Springer; 1995. p. 115–39.
- [12] Zehra Z, Luthra M, Siddiqui SM, Shamsi A, Gaur NA, Islam A. Corona virus versus existence of human on the earth: a computational and biophysical approach. *Int J Biol Macromol* 2020;vol. 161:271–81.
- [13] Shang W, Yang Y, Rao Y, Rao X. The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. *npj Vaccin* 2020;vol. 5(1):1–3.
- [14] W. Wodarg, Dangerous side effects of genetically induced production of SARS CoV-2 spike proteins.
- [15] Mohapatra S, et al. The novel SARS-CoV-2 pandemic: possible environmental transmission, detection, persistence and fate during wastewater and water treatment. *Sci Total Environ* 2021;vol. 765:142746.
- [16] Rayan RA, Zafar I, Tsagkaris C. Artificial Intelligence and Big Data Solutions for COVID-19. in *Intelligent Data Analysis for COVID-19 Pandemic*. Springer; 2021. p. 115–27.
- [17] COVID C, Team R. SARS-CoV-2 B. 1.1. 529 (Omicron) Variant—United States, December 1–8, 2021. *Morb Mortal Wkly Rep* 2021;vol. 70(50):1731.
- [18] Choudhary OP, Manish Dhawan P. Omicron variant (B. 1.1. 529) of SARS-CoV-2: Threat assessment and plan of action. *Int J Surg (Lond, Engl)* 2022;vol. 97:106187.
- [19] Jansen L. Investigation of a SARS-CoV-2 B. 1.1. 529 (Omicron) Variant Cluster—Nebraska, November–December 2021. *Mmwr Morb Mortal Wkly Rep* 2021;vol. 70.
- [20] Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX 1–2.3. *22 Curr Protoc Bioinforma* 2003;no. 1:2.3. 1–2.3. 22.
- [21] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;vol. 33(7):1870–4.
- [22] Wang J, et al. High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. *Mol Ecol* 2017;vol. 26(4):1091–107.
- [23] Peng B, Amos CI. Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics* 2008;vol. 24(11):1408–9.
- [24] R.C. Team, R: A language and environment for statistical computing, 2013.
- [25] Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2014;vol. 2:e281.

- [26] Paradis E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 2010;vol. 26(3):419–20.
- [27] Adamack AT, Gruber B. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol* 2014;vol. 5(4):384–7.
- [28] Brookfield J. A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* 1996;vol. 5(3):453–5.
- [29] Kalinowski ST. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conserv Genet* 2004;vol. 5(4):539–43.
- [30] Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol Evol* 2013;vol. 4(8):782–8.
- [31] Nei M. Genetic distance between populations. *Am Nat* 1972;vol. 106(949):283–92.
- [32] Marei HE, et al. Pandemic COVID-19 caused by SARS-CoV-2: genetic structure, vaccination, and therapeutic approaches. *Mol Biol Rep* 2021;vol. 48(9):6513–24.
- [33] Hatmal M m M, et al. Comprehensive structural and molecular comparison of spike proteins of SARS-CoV-2, SARS-CoV and MERS-CoV, and their interactions with ACE2. *Cells* 2020;vol. 9(12):2638.
- [34] Wei C, Shan K-J, Wang W, Zhang S, Huan Q, Qian W. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *J Genet Genom* 2021.
- [35] Zafar I, Pervez MT, Rather MA, Babar ME, Raza MA, Iftikhar R. Genome-wide identification and expression analysis of PPOs and POX gene families in the selected plant species. *Biosci Biotechnol Res Asia* 2020;vol. 17(2):301–18.
- [36] Kasibhatla SM, Kinikar M, Limaye S, Kale MM, Kulkarni-Kale U. Understanding evolution of SARS-CoV-2: a perspective from analysis of genetic diversity of RdRp gene. *J Med Virol* 2020;vol. 92(10):1932–7.
- [37] Motayo BO, et al. Evolution and genetic diversity of SARS-CoV-2 in Africa using whole genome sequences. *Int J Infect Dis* 2021;vol. 103:282–7.
- [38] Mengoni A, Bazzicalupo M. The statistical treatment of data and the analysis of molecular variance (AMOVA) in molecular microbial ecology. *Ann Microbiol* 2002;vol. 52(2):95–102.
- [39] Durmaz V, et al. Structural-bioinformatics analysis of SARS-CoV-2 variants reveals higher hACE2 receptor binding affinity for Omicron B. 1.1. 529 spike RBD Comp wild-Type Ref 2021.
- [40] Meraz M, Vernon-Carter E, Rodriguez E, Alvarez-Ramirez J. A fractal scaling analysis of the SARS-CoV-2 genome sequence. *Biomed Signal Process Control* 2022;vol. 73:103433.
- [41] Lan C, Kui-biao L, Yi-yun C. Whole-genome sequencing-based genetic analysis of influenza A (H3N2) virus strains isolated in Guangzhou city. *中国公共卫生* 2022;vol. 38(3):46–52.
- [42] Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* 2018;vol. 27(1):135–45.
- [43] Kandeel M, Mohamed ME, Abd El-Lateef HM, Venugopala KN, El-Beltagi HS. Omicron variant genome evolution and phylogenetics. *J Med Virol* 2022;vol. 94(4):1627–32.
- [44] Zafar I, et al. Genome-wide identification and analysis of GRF (growth-regulating factor) gene family in *Camila sativa* through in silico approaches. *J King Saud Univ-Sci* 2022;vol. 34(4):102038.
- [45] Zafar I, Iftikhar R, Ahmad SU, Rather MA. Genome wide identification, phylogeny, and synteny analysis of sox gene family in common carp (*Cyprinus carpio*). *Biotechnol Rep* 2021;vol. 30:e00607.
- [46] He L, Sun S, Zhang Q, Bao X, Li PK. Alignment-free sequence comparison for virus genomes based on location correlation coefficient. *Infect, Genet Evol* 2021;vol. 96:105106.
- [47] Peacock TP, et al. The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry 12. 31.474653 *BioRxiv*2022:2021. 12. 31.474653.
- [48] Sabir DK. Analysis of SARS-COV2 spike protein variants among Iraqi isolates. *Gene Rep* 2022;vol. 26:101420.
- [49] Mlcochova P, et al. SARS-CoV-2 B. 1.617. 2 Delta Var Émerv Vaccin Breakthr 2021.
- [50] Madhi SA, et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B. 1.351 variant. *N Engl J Med* 2021;vol. 384(20):1885–98.
- [51] Cameron E, et al. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* 2022;vol. 602(7898):664–70.
- [52] Liu L, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* 2022;vol. 602(7898):676–81.
- [53] Boni MF, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* 2020;vol. 5(11):1408–17.
- [54] Jin X, et al. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* 2020;vol. 69(6):1002–9.