

RESEARCH ARTICLE

Whole genome sequencing of SARS-CoV2 strains circulating in Iran during five waves of pandemic

Jila Yavarian¹, Ahmad Nejati¹, Vahid Salimi¹, Nazanin Zahra Shafiei Jandaghi¹, Kaveh Sadeghi¹, Adel Abedi², Ali Sharifi Zarchi³, Mohammad Mehdi Gouya⁴, Talat Mokhtari-Azad^{1*}

1 Virology Department, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, **2** Mathematics Department, Shahid Beheshti University, Tehran, Iran, **3** Department of Computer Engineering, Sharif University of Technology, Tehran, Iran, **4** Iranian Center for Communicable Disease Control, Tehran, Iran

* mokhtari@tums.ac.ir



OPEN ACCESS

Citation: Yavarian J, Nejati A, Salimi V, Shafiei Jandaghi NZ, Sadeghi K, Abedi A, et al. (2022) Whole genome sequencing of SARS-CoV2 strains circulating in Iran during five waves of pandemic. PLoS ONE 17(5): e0267847. <https://doi.org/10.1371/journal.pone.0267847>

Editor: Etsuro Ito, Waseda University: Waseda Daigaku, JAPAN

Received: December 14, 2021

Accepted: April 14, 2022

Published: May 2, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0267847>

Copyright: © 2022 Yavarian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All submitted files are available from the GISAID database (Accession numbers: EPI-ISL-1014676-87, EPI-ISL-959275-84, EPI-ISL-862075-81, EPI-ISL-1993547-557,

Abstract

Purpose

Whole genome sequencing of SARS-CoV2 is important to find useful information about the viral lineages, variants of interests and variants of concern. As there are not enough data about the circulating SARS-CoV2 variants in Iran, we sequenced 54 SARS-CoV2 genomes during the 5 waves of pandemic in Iran.

Methods

After viral RNA extraction from clinical samples collected during the COVID-19 pandemic, next generation sequencing was performed using the Nextseq platform. The sequencing data were analyzed and compared with reference sequences.

Results

During the 1st wave, V and L clades were detected. The second wave was recognized by G, GH and GR clades. Circulating clades during the 3rd wave were GH and GR. In the fourth wave GRY (alpha variant), GK (delta variant) and one GH clade (beta variant) were detected. All viruses in the fifth wave were in clade GK (delta variant). There were different mutations in all parts of the genomes but Spike-D614G, NSP12-P323L, N-R203K and N-G204R were the most frequent mutants in these studied viruses.

Conclusions

These findings display the significance of SARS-CoV2 monitoring to help on time detection of possible variants for pandemic control and vaccination plans.

EPI-ISL-2360250-57, EPI-ISL-4803556-58, EPI-ISL-4803554, EPI-ISL-4803538, and EPI-ISL-4803528).

Funding: A part of this study is supported by National Institute for Medical Research Development (NIMAD) under grant number 994376. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

In December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) was detected in Wuhan, China which has known as the cause of coronavirus disease 2019 (COVID-19) pandemic [<https://covid19.who.int/>].

To date, Iran has experienced five waves of the pandemic since the first detection of SARS-CoV2 on 19 February 2020 in Iran [1]. As of 16 November 2021, 253,640,693 cases of SARS-CoV2 with 5,104,899 deaths were reported worldwide and in Iran 6,045,212 laboratory confirmed cases with 128,272 deaths reported [<https://covid19.who.int/>].

SARS-CoV2 is a virus in *coronaviridae* family with a single stranded RNA. The genome is about 30 kb and consists of genes encoding multiple non-structural, structural and accessory proteins. The non-structural proteins include NSP1-16 which are necessary for virus transcription and replication. These proteins are encoded by ORF1ab with about 21300 nucleotides length [2].

NSP1 helps virus to evade innate host antiviral response and promote viral growth. NSP2 can bind to two host proteins prohibitin 1 and prohibitin 2 that are recognized to be important in cell migration, cell cycle progression, apoptosis, cellular differentiation, and mitochondrial biogenesis [3]. NSP3 is a papain-like protease and also it can suppress host protein synthesis. NSP3 and NSP4 with other cofactors are important for virus replication by prompting membrane rearrangement. NSP5 is a cysteine-like protease which is the virus main protease and cleaves NSP4-NSP16. NSP6 is involved in autophagy. NSP7 and NSP8 form complex with NSP12 for viral replicase machinery. NSP7 has primer-independent RNA polymerase activity. NSP8 has primase activity- Primase is an enzyme that synthesizes primers during replication. NSP9, an RNA-binding protein, in complex with NSP8, is involved in RNA replication and virulence. NSP10 is a cofactor for the 2' O-methyltransferase activity of NSP16 that increases evasion of the innate immune system, and the N7-guaninemethyltransferase/exoribonuclease activities of NSP14. NSP11 and NSP15 are involved in endoribonuclease activity and essential for replication. NSP12 has RNA polymerase activity. NSP13 has helicase and NTPase activity [2].

The structural proteins include spike (S), envelope (E), membrane (M) and nucleoprotein (N). Spike with a length of 1273 amino acid (aa), mediates attachment and entry. Envelope (75 aa) is a small membrane protein and important in virus infectivity. Membrane (222 aa) is important in virion morphogenesis and nucleoprotein (419 aa) is a viral genome packaging protein [2].

There are nine accessory proteins encoded by ORF3a, 3b, 6, 7a, 7b, 8, 9a, 9b and 10, which seem to be important in virus pathogenesis [2].

During virus replication, mutations can occur which might lead to alteration in protein functions and virus transmission and pathogenesis. Nowadays next-generation sequencing (NGS) is an effective method to identify different mutations and new variants for epidemiological and surveillance studies. SARS-CoV2 has different variants known as variants of concern (VOC), variants of interests (VOI) and variants under monitoring (VUM). It has also classified to different clades that currently 9 clades have been recognized based on markers mutations of the genome: S, L, V, G, and later of G into GH, GR and GV, and more recently GR into GRY and GRA [4]. Different variants of the SARS-CoV2 have been identified during the pandemic. Some spread worldwide, while others quickly faded away. Identification of the circulation of variants in a society is important, therefore, we set up NGS for SARS-CoV2 detected during five waves of pandemic for better understanding of circulation of different variants, genetic diversity and mutations in all non-structural, structural and accessory genes.

2 Materials and methods

2.1 Sample selection

Throat swab specimens from COVID-19 suspected patients were sent for sequencing to National Influenza Center (NIC) located at Virology Department, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. After primary detection by Real time PCR the samples from different parts of the country with ct value lower than 25 were selected for NGS. Iran experienced the 1st wave of SARS-CoV2 pandemic in February until May 2020. The 2nd wave began in late June until September 2020. The 3rd wave occurred during October to December 2020. The 4th wave started on early April until June 2021 and the 5th wave was from August until October 2021. This study was approved by ethics committee of National Institute for Medical Research Development, Tehran, Iran (IR.NIMAD.REC.1399.119).

2.2 Next generation sequencing (NGS)

After viral RNA extraction using High Pure Viral Nucleic Acid kit (Roche, Germany) according to the manufacturer's instruction, cDNA synthesis was performed. Library construction was done by using the Nextera DNA Flex kit (Illumina, USA), then used for hybridization using Respiratory Virus Oligo Panel kit (Illumina, USA). This was followed by bead-based capture of hybridized probes, amplification and clean-up. After quality control assessment for library concentration by Qubit (Thermo Fisher, USA), sequencing was performed using Next-Seq 550 machine (Illumina, USA).

2.3 Data & phylogenetic analysis

For data analysis, all collection of reads were mapped against the reference genome assembly of SARS-CoV2. We had high quality of assembled full viral genome coverage without undetermined nucleotides. The assembled genomes were analyzed using CoVsurver mutations App in GISAID [4] and aligned by BioEdit sequence alignment software. After sequence alignment, a phylogenetic tree was drawn by neighbor joining method (bootstrap 1000) using MEGA v7 software.

Then all 54 Iranian full genome SARS-CoV2 of different waves were submitted to GISAID with accession numbers: EPI-ISL-1014676-87, EPI-ISL-959275-84, EPI-ISL-862075-81, EPI-ISL-1993547-557, EPI-ISL-2360250-57, EPI-ISL-4803556-58, EPI-ISL-4803554, EPI-ISL-4803538 and EPI-ISL-4803528.

3 Results

Fifty four COVID-19 confirmed cases were subjected to NGS selecting 10, 10, 9, 20 and 5 samples from 1st, 2nd, 3rd, 4th and 5th waves respectively. Here we analyzed the mutations of each genes separately by comparing with hCoV-19/Wuhan/WIV04/2019 in GISAID.

3.1 Nonstructural proteins (ORF1ab)

The mutations of NSP genes are illustrated in the Fig 1. It is important to note that not all the mutations mentioned in the Fig 1 were found in all viruses during the related waves. Meanwhile in NSP10 and NSP11 no changes have been detected.

3.2 Structural proteins (S-E-M-N)

Spike. As shown in Table 1, amino acid substitutions in S glycoprotein were increased with pandemic progression. During the 1st wave there was no special amino acid substitutions

Table 1. Amino acid changes detected in spike glycoprotein during the five waves of SARS-CoV2 pandemic in Iran.

Sample name	Amino acid changes in Spike	Sample name	Amino acid changes in Spike
Wave 1			
hCoV-19/Iran/Tehran-08/2020		hCoV-19/Iran/Gilan-NIC189/2020	D614G
hCoV-19/Iran/Tehran-09/2020		hCoV-19/Iran/Ghilan-S15-319/2020	D614G
hCoV-19/Iran/Qom-907/2020		hCoV-19/Iran/Ghilan-T15-191/2020	D614G
hCoV-19/Iran/Gorgan-NIC992/2020		hCoV-19/Iran/Tehran-055M/2020	F1256L I210del L5F
hCoV-19/Iran/Gorgan-NIC118/2020		hCoV-19/Iran/Gilan-NIC184/2020	D614G
hCoV-19/Iran/Gorgan-NIC028/2020	G142S I210del	hCoV-19/Iran/Gilan-NIC230/2020	D614G
hCoV-19/Iran/Gorgan-NIC140/2020		hCoV-19/Iran/Ghilan-M30-241/2020	D614G
hCoV-19/Iran/Hamedan-66H/2020		hCoV-19/Iran/Gilan-NICS1-58/2020	D614G
hCoV-19/Iran/Urmia-715U/2020		hCoV-19/Iran/Ghilan-S1-32/2020	D614G
hCoV-19/Iran/Urmia-931U/2020	G142S I210del	hCoV-19/Iran/Qom-629/2020	D614G
Sample name			
Amino acid changes in Spike			
Wave 3			
hCoV-19/Iran/Tehran-NIC45RC/2020	D614G I210del		
hCoV-19/Iran/Tehran-NIC12RC/2020	D614G A845S	A846V D138Y	
hCoV-19/Iran/Tehran-04/2020	D614G Q314R		
hCoV-19/Iran/Tehran-06/2020	D614G	D138Y M177I	S477N
hCoV-19/Iran/Tehran-01/2020	D614G	D138Y	S477N
hCoV-19/Iran/Tehran-02/2020	D614G		
hCoV-19/Iran/Tehran-03/2020	D614G I210del		
hCoV-19/Iran/Tehran-05V/2020	D614G	D138Y	S477N
hCoV-19/Iran/Kashmar-15K/2020	D614G I210del		
Sample name			
Amino acid changes in Spike			
Wave 4			
hCoV-19/Iran/Tehran-NIC-K23/2021	D614G A570D	D6118H H69del	V70del Y144del
hCoV-19/Iran/Hormozghan-NIC-15/2021	D614G A570D	D6118H H69del	V70del Y144del
hCoV-19/Iran/Tehran-NIC-24/2021	D614G A570D	D6118H H69del	V70del Y144del I100T L699I
hCoV-19/Iran/Tehran-NIC-V30/2021	D614G A570D	D6118H H69del	V70del Y144del
hCoV-19/Iran/Tehran-NIC-SH17/2021	D614G A570D	D6118H H69del	V70del Y144del
hCoV-19/Iran/Tehran-NIC-30/2021	D614G A570D	D6118H H69del	V70del Y144del
hCoV-19/Iran/Tehran-NIC-23K/2021	D614G A570D	D6118H H69del	V70del Y144del L699I

(Continued)

Table 1. (Continued)

Sample name	Amino acid changes in Spike					Sample name					Amino acid changes in Spike				
	D614G	A570D	D1118H	H69del	Spike	N501Y	P681H	S982A	T716I	V70del	Y144del	I100T	L699I		
hCoV-19/Iran/Tehran-NIC-28/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del	I100T	L699I		
hCoV-19/Iran/Tehran-NIC-31/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del				
hCoV-19/Iran/Tehran-NIC-17/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del	I100T	L699I		
hCoV-19/Iran/Tehran-NIC-15/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del	I100T	L699I		
hCoV-19/Iran/Kerman-NIC-K1/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del				
hCoV-19/Iran/Kerman-NIC-K2/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del	I100T	L699I		
hCoV-19/Iran/Tehran-NIC-S7/2021	D614G	A570D	D1118H	H69del		N501Y	P681H	S982A	T716I	V70del	Y144del	I100T	L699I K1191N		
hCoV-19/Iran/Hormozghan-NIC-B7/2021	D614G	A243del	A701V	D80A		D215G	E484K	K417N	L242del	L244del	N501Y				
hCoV-19/Iran/Boshehr-NIC-P10/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K				
hCoV-19/Iran/Boshehr-NIC-P11/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K				
hCoV-19/Iran/Yazd-NIC-Y3/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K	D1257F	F1256L		
hCoV-19/Iran/Maku-NIC-M5/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K	T95I	S255F		
hCoV-19/Iran/Yazd-NIC-Y6/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K	E1202Q	L938F		
Wave 5															
hCoV-19/Iran/Maragheh-NIC-S2/2021	D614G	D950N						P681R		T19R	T478K	T95I	I850L		
hCoV-19/Iran/Ardakan-NIC-S5/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K	T95I	D574Y T299I		
hCoV-19/Iran/North Khorasan-NIC-S6/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K		T29A T250I		
hCoV-19/Iran/Ahvaz-NIC-S8/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K	T95I	A262S I850L		
hCoV-19/Iran/Ahvaz-NIC-S9/2021	D614G	D950N	E156G	F157del		G142D	L452R	P681R	R158del	T19R	T478K	T95I	A262S I850L		

<https://doi.org/10.1371/journal.pone.0267847.t001>

ORF lab																	
NSP 1	NSP2	NSP3			NSP4	NSP5	NSP6	NS P7	NSP 8	NSP 9	NSP 10	NSP 11	NSP12	NSP13	NSP14	NSP15	NSP16
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
M85del V86del	-	S135	G98V	-													
					1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave								
					R27C V198I	V198I S305I A476V	T160I A510V	K81N A375V E457K	S543P								
NSP3																	
1 ST wave	2 nd wave	3 rd wave			4 th wave			5 th wave									
S1670F	L689F T942I S1670F	T64I, S126L, E382K M494I, G773R, P874S, Q995H A1033V, A1798V, 182TTIQT186del			T183I, G282V, A488S, T725I S794L, K837N, A80D, T942K L1226P, P1228L, S1285F, I1412T P1469S, A1775V			T6I, S126L, A488S T1180I, P1228L P1469S, T1489M S1717L									
NSP4																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
M33I	-	T354A	V167L T492I	V167L T295I T492I													
NSP5																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
D248Y	P96S	-	V86I, K90R G283S	-													
NSP6																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
L37F	L37F Y80C	A79V	A46V, T77A L260F, 106SGF108del	T77A													
NSP7																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
-	E50G	N78S	-	-													
NSP8																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
T187I	I185V	-	-	K90R, T124I													
NSP9																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
-	-	-	-	Q49K													
NSP12																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
-	P323L, A400S, L638F	G137C, P323L, L638F	D284E, P323L, G671S	P323L, G671S													
NSP13																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
P78S, F475V	L176F, H290Y A302S, S485L	H164Y A598V	P77L S259L	P77L A598S													
NSP14																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
-	--	-	S147G A394V	V120I A394V													
NSP15																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
-	-	T112I	-	A6S, T325I													
NSP16																	
1 ST wave	2 nd wave	3 rd wave	4 th wave	5 th wave													
-	-	T91M	-	K160, L183F													

Fig 1. The substitutions of nonstructural proteins during the five waves of SARS-CoV2 pandemic in Iran.

<https://doi.org/10.1371/journal.pone.0267847.g001>

on the S glycoprotein but viruses detected during the 4th and 5th waves had the most frequent amino acid substitutions. It should be noted that D614G was detected in all viruses after the 1st wave.

Envelope. No mutation was found in viruses studied during the 1st, 2nd, 3rd and the 5th waves but in two samples in the 4th wave there were mutations encoding the amino acid

substitution P71L in one sample belonged to beta variant and S68F in one sample belonged to alpha variant.

Membrane. During the 1st and the 2nd waves, there were no mutations in M, but a mutation causing the amino acid substitution I73M was detected in three samples during the 3rd wave. Four samples during the 4th wave and 4 samples of the 5th wave had I82T which all these 8 samples belonged to delta variant.

Nucleoprotein. During the 1st wave, one sample had N192K in N and two samples had A35V. During the 2nd wave, two samples had D3Y and M234I, three samples had S194L substitutions. One sample had P13T and S194L. Another sample had A220V and P326L. hCoV-19/Iran/Gilan-NICS1-58/2020 along with four viruses during the 3rd wave, had R203K and G204R. hCoV-19/Iran/Tehran-05V/2020 had P162S beside R203K and G204R. One sample during the 3rd wave, had S201I. Three samples had S194L along with another substitution including G34W, P383L and P13T. During the 4th wave, 13 samples of alpha variant had D3L, R203K, G204R and S235F. One sample of alpha variant beside these four substitutions had also L219F. Seven samples belonged to delta variant had D63G, D377Y, G215C and R203M. One sample of delta variant during the 5th wave had D63G, R203M and D377Y. One sample had just R203M and D377Y and one sample had D63G, R203M and G215C. The only beta variant had T205I and T362I.

3.3 Accessory proteins

First wave. One sample had T151I in NS3, another sample had A105V in NS7a and two samples had T40I in NS7b.

Second wave. In NS3, five samples had Q57H and one sample had G18V along with Q57H. One sample had G254V and one sample had T223A and V112F, therefore during the 2nd wave eight samples had just mutation in NS3 among all accessory proteins.

Third wave. In NS3, two samples had Q57H the same as five samples in the 2nd wave. Two more samples had one more substitution along with Q57H in NS3, including W131C and T223I. One sample had just N58G in NS3. Three samples had a deletion at G66 (G66del) and S67 (S67del) and K68E in NS8. One sample had T223I and L4F in NS3.

Fourth wave. During the fourth wave, amino acid substitutions detected in NS3 were W131C, S171L, G100C, D238Y, S26L, Q57H and S171L in different samples. There was just W27L in NS6 in one sample. T39I, T120I, V82A and S83L were detected in NS7a. NS7b had T40I and a stop codon at E39 (E39stop). NS8 had the most frequent amino acid substitutions in viruses circulating during the 4th wave including stop codons at K68 (K68stop) and Q27 (Q27stop) and R52I, V62L, Y73C, C90F. These stop codons at NS7b and NS8 may lead to producing the truncated proteins. During the 5th wave, amino acid substitutions were detected in NS3a, NS7a and NS7b which K16T, S26L and L41F in NS3a, L49P, A50D, D51H, T120I, V82A in NS7a and T40I in NS7b were found.

3.4 Phylogenetic analysis

Fig 2 shows the phylogenetic analysis of 54 SARS-CoV2 viruses circulating during the different waves in Iran. Analyses showed that during the 1st wave clades V and L were circulating which V was dominant. During the 2nd wave, GH was dominant, but G and GR were also detected. Surprisingly one V clade was also detected in the 2nd wave. GH and GR were the clades during the 3rd wave. During the 4th wave, 14 viruses of alpha variant were belonged to GRY, five delta variants were in GK clade and one beta variant was in GH clade and finally five delta variants of the 5th wave were in GK clade.

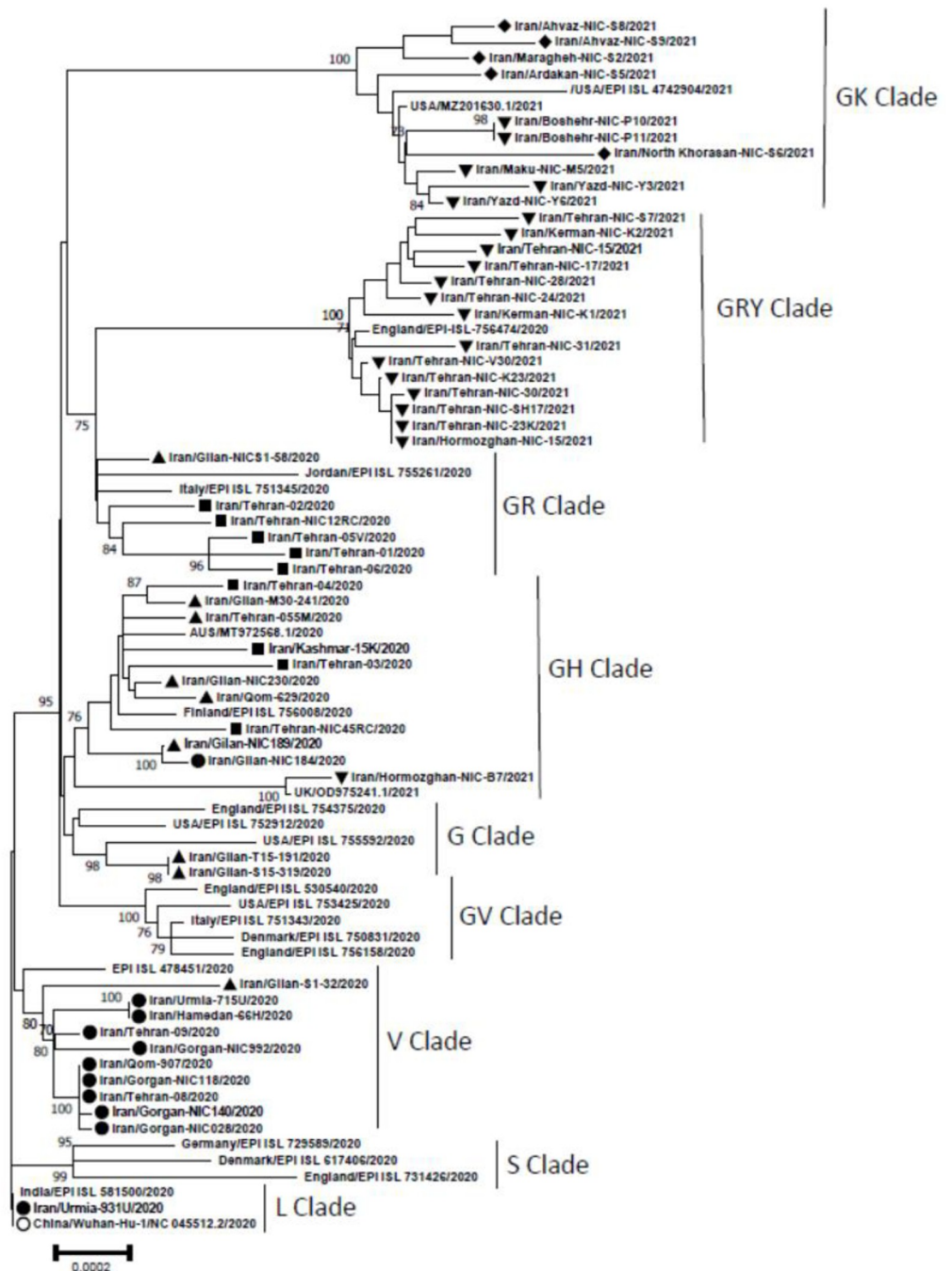


Fig 2. Phylogenetic tree of SARS-CoV2 full-length genomes constructed by MEGA 7. The Neighbor joining method was used with 1,000 bootstrap replicates. The tree contains 54 SARS-CoV2 sequences of these study compared to the reference sequence from GISAID and some other sequences from each clade. In this tree the reference sequence is marked by white circle and sequences of this study were marked as follow: The 1st wave black circle, the 2nd wave inverted black triangle, the 3rd wave with black square, the 4th wave with black triangle and the 5th wave with black diamond.

<https://doi.org/10.1371/journal.pone.0267847.g002>

4 Discussion

In this study, we reported the circulation of distinct clades of SARS-CoV2 during the five waves in Iran. The V clade was found during the 1st wave. Clade V is characterized by NSP6-L37F plus NS3-G251V. At the end of the first wave, one sample detected in clade L (reference clade). The 2nd wave was recognized by G, GH and GR clades. G clade shows D614G in spike which is one of the most important substitutions in spike. Clade GH has S-D614G plus NS3-Q57H substitutions and clade GR characterized by S-D614G and N-G204R. One sample belonged to the V clade was also detected in the 2nd wave. Circulating clades during the 3rd wave were GH and GR. During the fourth wave, GH (1 beta), GRY (alpha) and GK (delta) clades were detected. GRY clade (alpha) has S-H69del, S-V70del, S-Y144del, S-N501Y plus S-D614G and N-G204R. All viruses in the 5th wave were in clade GK (S-D614G plus S-T478K).

In different clades NGS of 54 samples selected during the five waves in Iran showed important mutations in the different parts of the genome which all were analyzed and compared with circulating variants worldwide. The effects of important mutations were discussed here.

ORF1ab is more than two thirds of the SARS-CoV2 genome with 21,290 nucleotides at the 5' end which encodes 16 non-structural proteins (NSP1-NSP16). Among these NSPs, NSP3 had the highest number of mutations. NSP3 is important for virus replication and it can suppress host protein synthesis, then amino acid substitutions in NS3 deserve greater study in vitro.

Suppression of type I interferon (IFN) response is a consequence of infection of SARS-CoV2. NSP5, main protease, is an IFN antagonist. NSP5 variant K90R seen in SARS-CoV2 retained the IFN-antagonizing activity. The suppressive outcome of NSP5 on IFN- β gene transcription induced by IKK ϵ , TBK1, RIG-I and MAVS suggested that NSP5 probably involved at a stage downstream of IRF3 phosphorylation in the cytoplasm [5]. In this study K90R was detected just in one sample during the 4th wave while it has been more often found in Icelandic and Chinese strains [6].

NSP6 is important for viral assembly, viral protein folding and replication. NSP6-L37F leads to asymptomatic transmission and reduced virulence which we can see in the most of the viruses detected during the 1st wave and one in the 2nd wave [7].

In hCoV-19/Iran/Gilan-NIC230/2020, E50G in NSP7 was detected in the 2nd wave which is important in immune response as shown in a study that amino acid residues 36–50 (HNDILLAKDTTEAFE) of NSP7 and also NSP13 and N are SARS-CoV2 specific T cell epitopes recognized by CD8 T cells [8].

NSP12 is necessary for the replication/transcription of the SARS-CoV2 genome, and this protein is considered as a target for the treatment of COVID-19. The P323L substitution of NSP12, could induce structural changes and adverse effect on proofreading during the replication of the virus. Meanwhile, the P323L is located in a pocket that might be the site of drug function [9]. In our research NSP12-P323L located in the NSP8 binding cleft [10] was detected after the first wave. NSP12-P323L is the most common detected substitution with increasing in occurrence over time [11].

A study by Mohammad., et al. showed that remdesivir has higher binding affinity to NSP12 with P323L than the wild-type RdRp, therefore they suggested using remdesivir in patients infected with SARS-CoV2 carrying P323L in NSP12 [12].

In this study there was no virus seen to contain F480, V557 and S861 in NSP12, which have shown to be important in decreasing susceptibility to remdesivir [11].

As all viruses studied in this research, did not have any mutations in the genes for NSP10 and NSP11, therefore these two nonstructural proteins could be considered as good targets for

diagnosis and targets for therapeutic agents. Furthermore, there were some mutations in other NSPs with that of unknown importance. The effect of these substitutions need to be studied in vitro.

After the 1st wave, all viruses studied in this research had S-D614G. Studies showed that the S-protein amino acid substitution, D614G, evolved concurrently with P323L in NSP12. NSP12- P323L might decrease the replication of viral RNA and therefore increase the probability of asymptomatic infections and/or change in incubation period. Meanwhile the infectivity-enhancing D614G substitution in S protein could compensate this reduction in virus replication and help virus transmission even by asymptomatic individuals [13]. Flores-Alanis, et al. showed that D614G substitution in S protein, P323L in NSP12, and R203K and G204R in N protein had a substantial association with the disease severity [9]. Another study showed that D614G in the S protein increased the infectivity of SARS-CoV2 and associated with lower RT-PCR cycle thresholds. It showed high viral load in the upper respiratory tract, but not enhanced disease severity. Also they showed that G614-bearing virions are not intrinsically more resistant to neutralization by convalescent sera [14]. G614 of S protein is responsible for making the virus 2.4 times more contagious with higher viral load [9, 15]. hCoV-19/Iran/Tehran-055M/2020 in the 2nd wave had D614G and L5F. One study showed that variants with D614G and L5F had increased infectivity [16].

Evaluation of the mutations which lead to amino acid substitutions, showed that Spike-D614G, NSP12-P323L, N-R203K and N-G204R were the most frequent amino acid substitutions in these studied viruses and also worldwide which we could consider that they may increase SARS-CoV2 transmissibility as the pandemic progressed.

The S protein is crucial factor for the entry of SARS-CoV2 to the host cell which interacts with the angiotensin-converting enzyme 2 (ACE-2) receptor through its receptor binding domain (RBD) [17]. The S477N is the part of an epitope recognized by human neutralizing antibodies and located in the RBD. A study showed that S477N increases the affinity for the ACE-2 receptor [9, 18]. Singh, et al. showed that S477N strengthen the binding of SARS-CoV2 spike to the ACE-2 receptor [19]. In this study, two samples had S477N which both belonged to GR clade.

Among the variations in the alpha variant, S-N501Y, is in the receptor binding site which was shown to increase the binding of SARS-CoV2 to the ACE-2 host receptor, leading to increased viral fitness and transmission [20–22].

One study showed that N439K, N501Y and S477N significantly reduced the neutralization activity of some monoclonal antibodies [23]. Ostrov showed that B.1.1.7 with N501Y had increased affinity for ACE-2 and A570D, D614G and S982A substitutions might increase virus fusion by decreasing the intermolecular stability of S1 and S2 [24].

The delta variant (GK clade) has ten amino acid substitutions, T19R, G142D, 156del, 157del, R158G, L452R, T478K, D614G, P681R and D950N in the spike protein [25]. T19R removes an N-glycosylation site at position 17 that might also affect antigenic properties. A study by Li, et al. showed that variants with L452R were resistant to some neutralizing antibodies [16]. Tchesnokova, et al. showed that L452R might result in stronger binding to the ACE-2 and escape from neutralizing antibodies [26].

Overall L452R reduces sensitivity to neutralizing antibodies, increases viral infectivity, transmissibility, spike stability, ACE-2 binding affinity and viral fusogenicity, therefore it supports viral replication [27–30].

Delta variant circulation was started at the end of the 4th wave and all viruses detected during the 5th wave in Iran were the delta variant.

The beta variant (GH clade) includes L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, and A701V in the spike protein, which K417N, E484K, and N501Y are in the RBD

and increase the binding affinity for the ACE-2 receptor and eventually the infectivity [31]. K417N and E484K have also essential role in viral escape from neutralizing antibodies [32]. Overall K417N, E484K and L452R are vaccine escape mutants [33]. The only beta variant of this study did not have L18F and R246I but it had a deletion of three amino acids 242–244 (L242del, A243del and L244del) in spike. L242del, A243del and L244del showed reduced sensitivity to some neutralizing antibodies [22]. The beta variant is reported to have an increased risk of transmission and reduced neutralization by monoclonal antibody therapy, convalescent sera, and post-vaccination sera [34].

Some studies showed that most serum samples from vaccinated people or patients recovering from COVID-19 have shown full or slightly decreased capacity to inactivate SARS-CoV2 variants, except for variants with N501Y, K417N and E484K substitutions [22, 35–37]. We had just one beta variant during the 4th wave which had E484K in the S protein. As seen in different studies, E484K plays a crucial role in increasing virus transmission and decreasing antibody neutralizing titers [37–39], thus continuous screening for emerging variants with substitutions such as E484K is necessary for public health.

hCoV-19/Iran/Tehran-03/2020 in the 3rd wave had Q677H. Grabowski, et al. showed that E484K, F490S, S494P (in the RBD of spike) and Q677H and Q675H (in the vicinity of the polybasic cleavage site at the S1/S2 border) may limit efficiency of vaccines [40].

Liu, et al. showed that substitutions at residues T345, R346, K444, G446, N450, L452, S477, T478, E484, F486, and P499 were each related to the resistance to more than one monoclonal antibody, of which substitutions at S477 and E484 residues showed wide resistance [41].

Viruses with H69del and V70del had 2-fold higher infectivity compared to wildtype and also showed reduced neutralization sensitivity to mAb, targeting an as yet undefined epitope outside the RBD [42]. Sixteen samples of this study during the 4th wave had H69del, V70del and also Y144del. In one study it was shown that viruses with Y144del in spike had decreased sensitivity to convalescent sera [16].

Li, et al. showed that, among spike mutations, the most characteristic ones are substitutions such as D614G, N501Y, Y453F, N439K/R, P681H, K417N/T, and E484K, and deletions of Δ H69/V70 and Δ 242–244, which enhanced viral infectivity, transmissibility, and resistance to neutralization [43]. Some of these mutations were found after the 2nd wave of pandemic in Iran.

P681R in the furin cleavage site may help in increased rate of membrane fusion, internalization and so better transmissibility which was found in delta variants in this study and worldwide [44]. Voss, et al. found that P681H of spike and S235F of nucleoprotein in the alpha variant changed the specificity of the corresponding epitopes [45]. During the 4th wave, 14 samples had P681H and S235F.

Mutations in the E gene were somewhat uncommon as in this study just two mutations in 2 different samples in the 4th wave were detected which needs more investigation to find out their effects on virus life cycle.

All viruses of delta variant in this study had I82T in M gene. Shen, et al. showed that this M gene substitution was more naturally fit, probably connected to glucose uptake during virus replication, and it is better to be included in genomic surveillance [25].

Nucleoprotein was mainly expressed in the initial stages of infection, and is important in viral RNA transcription and replication. Nucleoprotein has been shown to affect some basic cellular processes, inflammatory responses to upregulate the expression of the proinflammatory factor COX2, and it inhibits the innate immune responses in the host cell [46]. Therefore, amino acid substitutions in the nucleoprotein might have significant effect in immune response. R203K and G204R were the most common substitutions in the nucleoprotein in

viruses studied in this research and worldwide. These mutations are important in disease severity [47].

There are some studies about the accessory gene's mutations and their impact on the virus cell cycle. The results of research by Wu et al., showed that Q57H and G251V in NS3a, S194L and R203K/G204R in N made changes in the structure of proteins and also had effect on the binding affinity of intraviral protein-protein interactions during assembly and release of coronavirus. So these changes might be associated with virus evolution and beneficial for the virus and its virulence [48]. Some viruses during the waves 2,3 and 4 had NS3a-Q57H. In primary human respiratory cells, viruses with NS3a-Q57H evade stimulation of chemokine, cytokine, and interferon-stimulated gene expression [49]. Some studies have described that NS3a can competently induce apoptosis in the cell and affect virus uptake and release [50, 51]. Besides ORF3, some mutations were detected in different accessory genes of viruses in this study which more research is needed to explore their importance in virus pathogenesis. For instance, NS8 supposed to suppress immune responses [52], then substitutions of this protein can be imperative. Of special importance, stop codon mutations which lead to absence of NS8 can extend the duration of symptoms, then increase the virus transmissibility. Spike mutations such as D614G, HV69-70 del and L5F which affect the receptor binding affinity of the S protein and increase the virus transmission when associated with NS8 stop codons which extend the period of signs and symptoms might enhance the chance of transmission [53]. The combination of these variations were found especially in some viruses detected during the 4th wave which should raise concern.

In this study we did not have data on relationship between different SARS-CoV2 lineages and patient's signs and symptoms and disease severity which was an important limitation of this research.

In conclusion, we detected different lineages of SARS-CoV2 contributing to all five waves and showed that all viruses circulating during the 5th wave belonged to delta variant. We compared the mutations identified in our complete genomes study with those reported in GISAID. The findings of this study showed that with progression of the pandemic, the number of mutations were considerably increased which showed the adaptive evolution of SARS-CoV2 in human to increase transmissibility. Therefore genomic surveillance is an important tool to screen the progression of the COVID-19 pandemic. It should be noted that for variant detection, we partially sequenced S glycoprotein of more than 1000 samples with Sanger sequencer of which the results were compatible with NGS results during all 5 waves (unpublished data).

Acknowledgments

Authors would like to express their great thanks to all data contributors, i.e. the authors and their originating laboratories responsible for obtaining the specimens, and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We thank GISAID for all their support.

Author Contributions

Data curation: Adel Abedi.

Formal analysis: Ali Sharifi Zarchi.

Methodology: Ahmad Nejati, Vahid Salimi, Nazanin Zahra Shafiei Jandaghi, Kaveh Sadeghi.

Supervision: Mohammad Mehdi Gouya, Talat Mokhtari-Azad.

Writing – original draft: Jila Yavarian.

References

1. Yavarian J, Shafiei-Jandaghi N-Z, Sadeghi K, Malekshahi SS, Salimi V, Nejati A, et al. First cases of SARS-CoV-2 in Iran, 2020: case series report. *Iranian Journal of Public Health*. 2020; 49(8):1564. <https://doi.org/10.18502/ijph.v49i8.3903> PMID: 33083334
2. Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, Dhamija P, et al. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*. 2021; 10(4):821. <https://doi.org/10.3390/cells10040821> PMID: 33917481
3. Cornillez-Ty CT, Liao L, Yates JR III, Kuhn P, Buchmeier MJ. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *Journal of virology*. 2009; 83(19):10314–8. <https://doi.org/10.1128/JVI.00842-09> PMID: 19640993
4. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Euro-surveillance*. 2017; 22(13):30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494> PMID: 28382917
5. Fung S-Y, Siu K-L, Lin H, Yeung ML, Jin D-Y. SARS-CoV-2 main protease suppresses type I interferon production by preventing nuclear translocation of phosphorylated IRF3. *International Journal of Biological Sciences*. 2021; 17(6):1547. <https://doi.org/10.7150/ijbs.59943> PMID: 33907518
6. Yuan F, Wang L, Fang Y, Wang L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. *Transboundary and emerging diseases*. 2020. <https://doi.org/10.1111/tbed.13931> PMID: 33207070
7. Wang R, Chen J, Hozumi Y, Yin C, Wei G-W. Decoding asymptomatic COVID-19 infection and transmission. *The journal of physical chemistry letters*. 2020; 11(23):10007–15. <https://doi.org/10.1021/acs.jpcclett.0c02765> PMID: 33179934
8. Le Bert N, Tan AT, Kunasegaran K, Tham CY, Hafezi M, Chia A, et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature*. 2020; 584(7821):457–62. <https://doi.org/10.1038/s41586-020-2550-z> PMID: 32668444
9. Flores-Alanis A, Cruz-Rangel A, Rodríguez-Gómez F, González J, Torres-Guerrero CA, Delgado G, et al. Molecular epidemiology surveillance of SARS-CoV-2: mutations and genetic diversity one year after emerging. *Pathogens*. 2021; 10(2):184. <https://doi.org/10.3390/pathogens10020184> PMID: 33572190
10. Mutlu O, Ugurel OM, Sariyer E, Ata O, Inci TG, Ugurel E, et al. Targeting SARS-CoV-2 Nsp12/Nsp8 interaction interface with approved and investigational drugs: an in silico structure-based approach. *Journal of Biomolecular Structure and Dynamics*. 2020:1–13.
11. Martin R, Li J, Parvangada A, Perry J, Cihlar T, Mo H, et al. Genetic conservation of SARS-CoV-2 RNA replication complex in globally circulating isolates and recently emerged variants from humans and minks suggests minimal pre-existing resistance to remdesivir. *Antiviral research*. 2021; 188:105033. <https://doi.org/10.1016/j.antiviral.2021.105033> PMID: 33549572
12. Mohammad A, Al-Mulla F, Wei D-Q, Abubaker J. Remdesivir MD Simulations Suggest a More Favourable Binding to SARS-CoV-2 RNA Dependent RNA Polymerase Mutant P323L Than Wild-Type. *Biomolecules*. 2021; 11(7):919. <https://doi.org/10.3390/biom11070919> PMID: 34206274
13. Ilmjärvi S, Abdul F, Acosta-Gutiérrez S, Estarellas C, Galdadas I, Casimir M, et al. Concurrent mutations in RNA-dependent RNA polymerase and spike protein emerged as the epidemiologically most successful SARS-CoV-2 variant. *Scientific reports*. 2021; 11(1):1–13.
14. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. 2020; 182(4):812–27. e19. <https://doi.org/10.1016/j.cell.2020.06.043> PMID: 32697968
15. Vilar S, Isom DG. One year of SARS-CoV-2: How much has the virus changed? *Biology*. 2021; 10(2):91. <https://doi.org/10.3390/biology10020091> PMID: 33530355
16. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*. 2020; 182(5):1284–94. e9. <https://doi.org/10.1016/j.cell.2020.07.012> PMID: 32730807
17. Shatzadeh Malekshahi S, Yavarian J, Shafiei-Jandaghi NZ. Usage of peptidases by SARS-CoV-2 and several human coronaviruses as receptors: A mysterious story. *Biotechnology and Applied Biochemistry*. 2020.
18. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KH, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*. 2020; 182(5):1295–310. e20. <https://doi.org/10.1016/j.cell.2020.08.012> PMID: 32841599
19. Singh A, Steinkellner G, Köchl K, Gruber K, Gruber CC. Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. *Scientific reports*. 2021; 11(1):1–11.

20. Mallon PW, Crispie F, Gonzalez G, Tinago W, Leon AG, McCabe M, et al. Whole-genome sequencing of SARS-CoV-2 in the Republic of Ireland during waves 1 and 2 of the pandemic. medRxiv. 2021.
21. Rachiglio AM, De Sabato L, Roma C, Cennamo M, Fiorenza M, Terracciano D, et al. SARS-CoV-2 complete genome sequencing from the Italian Campania region using a highly automated next generation sequencing system. *Journal of Translational Medicine*. 2021; 19(1):1–10.
22. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al. SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma. *Nature medicine*. 2021; 27(4):622–5. <https://doi.org/10.1038/s41591-021-01285-x> PMID: 33654292
23. Wu J, Zhang L, Zhang Y, Wang H, Ding R, Nie J, et al. The Antigenicity of Epidemic SARS-CoV-2 Variants in the United Kingdom. *Frontiers in Immunology*. 2021; 12:2205. <https://doi.org/10.3389/fimmu.2021.687869> PMID: 34220844
24. Ostrov DA. Structural Consequences of Variation in SARS-CoV-2 B. 1.1. 7. *Journal of cellular immunology*. 2021; 3(2):103. <https://doi.org/10.33696/immunology.3.085> PMID: 33969357
25. Shen L, Bard JD, Triche TJ, Judkins AR, Biegel JA, Gai X. Emerging variants of concern in SARS-CoV-2 membrane protein: a highly conserved target with potential pathological and therapeutic implications. *Emerging microbes & infections*. 2021; 10(1):885–93. <https://doi.org/10.1080/22221751.2021.1922097> PMID: 33896413
26. Tchesnokova V, Kulakesara H, Larson L, Bowers V, Rechkina E, Kisiela D, et al. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. *BioRxiv*. 2021.
27. Deshpande A, Harris BD, Martinez-Sobrido L, Kobie JJ, Walter MR. Epitope classification and RBD binding properties of neutralizing antibodies against SARS-CoV-2 variants of concern. *Frontiers in immunology*. 2021; 12:2185. <https://doi.org/10.3389/fimmu.2021.691715> PMID: 34149735
28. Motozono C, Toyoda M, Zahradnik J, Saito A, Nasser H, Tan TS, et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell host & microbe*. 2021; 29(7):1124–36. e11. <https://doi.org/10.1016/j.chom.2021.06.006> PMID: 34171266
29. Díaz Y, Ortiz A, Weeden A, Castillo D, González C, Moreno B, et al. SARS-CoV-2 reinfection with a virus harboring mutation in the Spike and the Nucleocapsid proteins in Panama. *International Journal of Infectious Diseases*. 2021. <https://doi.org/10.1016/j.ijid.2021.06.004> PMID: 34107326
30. Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell*. 2021.
31. Mohammadi M, Shayestehpour M, Mirzaei H. The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines. *Brazilian Journal of Infectious Diseases*. 2021; 25.
32. Moyo-Gwete T, Madzivhandila M, Makhado Z, Ayres F, Mhlanga D, Oosthuysen B, et al. SARS-CoV-2 501Y. V2 (B. 1.351) elicits cross-reactive neutralizing antibodies. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.03.06.434193> PMID: 33688657
33. Wang R, Chen J, Gao K, Wei G-W. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries. *Genomics*. 2021; 113(4):2158–70. <https://doi.org/10.1016/j.ygeno.2021.05.006> PMID: 34004284
34. Yang W-T, Huang W-H, Liao T-L, Hsiao T-H, Chuang H-N, Liu P-Y. SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations. *Infection and Drug Resistance*. 2022; 15:373. <https://doi.org/10.2147/IDR.S344099> PMID: 35140483
35. Wadman M, Cohen J. Novavax vaccine delivers 89% efficacy against COVID-19 in UK—but is less potent in South Africa. *Science*. 2021.
36. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*. 2021; 592(7855):616–22. <https://doi.org/10.1038/s41586-021-03324-6> PMID: 33567448
37. Greaney AJ, Loes AN, Crawford KH, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell host & microbe*. 2021; 29(3):463–76. e6. <https://doi.org/10.1016/j.chom.2021.02.003> PMID: 33592168
38. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B. 1.351 and B. 1.1. 7. *Nature*. 2021; 593(7857):130–5. <https://doi.org/10.1038/s41586-021-03398-2> PMID: 33684923
39. West AP Jr, Wertheim JO, Wang JC, Vasylyeva TI, Havens JL, Chowdhury MA, et al. Detection and characterization of the SARS-CoV-2 lineage B. 1.526 in New York. *bioRxiv*. 2021.

40. Grabowski F, Preibisch G, Giziński S, Kocharczyk M, Lipniacki T. SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations. *Viruses*. 2021; 13(3):392. <https://doi.org/10.3390/v13030392> PMID: 33804556
41. Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell host & microbe*. 2021; 29(3):477–88. e4. <https://doi.org/10.1016/j.chom.2021.01.014> PMID: 33535027
42. Kemp SA, Collier DA, Datir RP, Ferreira IA, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021; 592(7853):277–82. <https://doi.org/10.1038/s41586-021-03291-y> PMID: 33545711
43. Li X, Zhang L, Chen S, Ji W, Li C, Ren L. Recent progress on the mutations of SARS-CoV-2 spike protein and suggestions for prevention and controlling of the pandemic. *Infect Genet Evol*. 2021 Sep; 93:104971. <https://doi.org/10.1016/j.meegid.2021.104971> PMID: 34146731
44. Cherian S, Potdar V, Jadhav S, Yadav P, Gupta N, Das M, et al. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms*. 2021 Jul 20; 9(7). <https://doi.org/10.3390/microorganisms9071542> PMID: 34361977
45. Voss C, Esmail S, Liu X, Knauer MJ, Ackloo S, Kaneko T, et al. Epitope-specific antibody responses differentiate COVID-19 outcomes and variants of concern. *JCI Insight*. 2021 Jul 8; 6(13). <https://doi.org/10.1172/jci.insight.148855> PMID: 34081630
46. Surjit M, Liu B, Kumar P, Chow VT, Lal SK. The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. *Biochem Biophys Res Commun*. 2004 May 14; 317(4):1030–6. <https://doi.org/10.1016/j.bbrc.2004.03.154> PMID: 15094372
47. Wu H, Xing N, Meng K, Fu B, Xue W, Dong P, et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell host & microbe*. 2021; 29(12):1788–801. e6.
48. Wu S, Tian C, Liu P, Guo D, Zheng W, Huang X, et al. Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. *J Med Virol*. 2021 Apr; 93(4):2132–40. <https://doi.org/10.1002/jmv.26597> PMID: 33090512
49. Chu DKW, Hui KPY, Gu H, Ko RLW, Krishnan P, Ng DYM, et al. Introduction of ORF3a-Q57H SARS-CoV-2 Variant Causing Fourth Epidemic Wave of COVID-19, Hong Kong, China. *Emerg Infect Dis*. 2021 May; 27(5):1492–5. <https://doi.org/10.3201/eid2705.210015> PMID: 33900193
50. Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, et al. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cell Mol Immunol*. 2020 Aug; 17(8):881–3. <https://doi.org/10.1038/s41423-020-0485-9> PMID: 32555321
51. Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. *mSystems*. 2020 May 5; 5(3).
52. Flower TG, Buffalo CZ, Hooy RM, Allaire M, Ren X, Hurley JH. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proceedings of the National Academy of Sciences*. 2021; 118(2). <https://doi.org/10.1073/pnas.2021785118> PMID: 33361333
53. Pereira F. SARS-CoV-2 variants combining spike mutations and the absence of ORF8 may be more transmissible and require close monitoring. *Biochem Biophys Res Commun*. 2021 Apr 23; 550:8–14. <https://doi.org/10.1016/j.bbrc.2021.02.080> PMID: 33676232