Highlight of potential impact of new viral genotypes of SARS-CoV-2 on vaccines and anti-viral therapeutics

Abozar Ghorbani [a,*], Samira Samarfard [b], Maziar Jajarmi [c], Mahboube Bagheri [d], Thomas P. Karbanowicz [g], Alireza Afsharifar [a], Mohammad Hadi Eskandari [e], Ali Niazi [f], Keramatollah Izadpanah [a]

[a] Plant Virology Research Centre, College of Agriculture, Shiraz University, Shiraz, Iran
[b] Berrimah Veterinary Laboratory, Department of Primary Industry and Resources, Berrimah, NT 0828 Australia
[c] Department of Pathobiology, Faculty of Veterinary Medicine, Shahid Bahonar University of Kerman, Kerman, Iran
[d] Department of Food Science and Technology, Bardsir Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran
[e] Department of Food Science and Technology, College of Agriculture, Shiraz University, Shiraz, Iran
[f] Institute of Biotechnology, College of Agriculture, Shiraz University, Shiraz, Iran
[g] Australian Red Cross Lifeblood, West Melbourne 3003, VIC Australia

ARTICLE INFO

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causal agent of the coronavirus disease (COVID-19) pandemic, has infected millions of people globally. Genetic variation and selective pressures lead to the accumulation of single nucleotide polymorphism (SNP) within the viral genome that may affect virulence, transmission rate, viral recognition and the efficacy of prophylactic and interventional measures. To address these concerns at the genomic level, we assessed the phylogeny and SNPs of the SARS-CoV-2 mutant population collected to date in Iran in relation to globally reported variants. Phylogenetic analysis of mutant strains revealed the occurrence of the variants known as B.1.1.7 (Alpha), B.1.525 (Eta), and B.1.617 (Delta) that appear to have delineated independently in Iran. SNP analysis of the Iranian sequences revealed that the mutations were predominantly positioned within the S protein-coding region, with most SNPs localizing to the S1 subunit. Seventeen S1-localizing SNPs occurred in the RNA binding domain that interacts with ACE2 of the host cell. Importantly, many of these SNPs are predicted to influence the binding of antibodies and anti-viral therapeutics, indicating that the adaptive host response appears to be imposing a selective pressure that is driving the evolution of the virus in this closed population through enhancing virulence. The SNPs detected within these mutant cohorts are addressed with respect to current prophylactic measures and therapeutic interventions.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiological agent of the coronavirus disease (COVID-19) pandemic, has now infected more than 250 million (250,524,307) people and caused more than 5 million (5,059,893) deaths in one of the worst global pandemics of the recent century (Data accurate as of 9/11/2021: Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), COVID-19 Dashboard; https://coronavirus.jhu.edu/map.html). More than 6 million (6,012,408) COVID-19-infected cases and about 127, 439 cumulative death have been reported in Iran (https://covid19. who.int/region/emro/country/ir). Since the beginning of the pandemic, COVID-19-related policies have been implemented in all countries to restrict the spread of the virus and prevent the exhaustion of the health system's resources (Greer et al., 2020). Iran's health ministry has called for lockdowns and restrictions on entry to the high risks provinces

enforced by armed forces and law enforcement to limit the virus transmission and during the Persian new year (March 20 to May 2) (Hadianfar et al., 2021). However, only a few study cases are available on the efficiency of Iranian public health measures for reducing the rate of COVID-19-infected cases and the percentage of compliance or noncompliance to the control measures (Ghadiri et al., 2021; Wong et al., 2021).

About 59% of Iran's population (85 million) have received at least one vaccine dose and about 45% are fully vaccinated (Data accurate as of 3/11/2021: https://www.nytimes.com/2021/10/20/world/middleeast/iran-covid-vaccine-fakhravac.html).

The frequency of the viral mutations can be reduced through surging the rate of full vaccination and herd immunity. Therefore, countries with high vaccine coverage are less likely to experience the emergence of vaccine-resistant strains and new superspreading events (Rella et al., 2021).

Taxonomically, SARS-CoV-2 is a member of the *Coronaviridae* family, *Orthocoronavirinae* subfamily, and *Betacoronavirus* genus, which encompasses additional human pathogens including SARS-CoV and MERS-CoV. SARS-CoV-2 is an enveloped virus possessing a monopartite, positive-sense, single-stranded RNA genome consisting of 29,891 nucleotides that include two untranslated regions (UTRs) at the 5′ and 3′ ends and 12 putative Open Reading Frames (ORFs) in gene order from 5′ to 3′ that encode accessory proteins, non-structural proteins (NSP) and structural proteins (SP) (Feng et al., 2020; Harapan et al., 2020; Shaw et al., 2020). The 5′-terminus codes for ORF1a and ORF1b. The −1 ribosomal frameshift upstream of the ORF1a stop codon allows continued translation of the ORF1b coding region to generate a full-length ORF1ab polyprotein (Sola et al., 2015). The 3′-terminal ORFs of SARS-CoV-2 genome encode SPs, including spike glycoprotein (S, ORF2), envelope (E, ORF4), membrane (M, ORF5) and nucleocapsid (N, ORF9a) and accessory proteins (3a, 6, 7a, 7b, 8, and 10) that are expressed from nine predicted sub-genomic RNAs (Wu et al., 2020).

The surface glycoprotein (≈180 kDa) of SARS-CoV-2, known as S protein, is critical to viral attachment of ACE2 (angiotensin-converting enzyme 2), its cognate receptor on the surface of host cells derived from different vertebrate species (Jaimes et al., 2020a). The S protein of SARS-CoV-2 is composed of fusion peptide (FP), heptad repeat 1 (HR1), heptad repeat 2 (HR2), intracellular domain (IC), N-terminal domain (NTD), subdomain 1 (SD1), subdomain 2 (SD2), transmembrane region (TM), receptor-binding domain (RBD). In all coronaviruses including SARS-CoV-2, the S-glycoprotein is cleaved by host proteases at the S1/S2 junction. This cleavage activates S protein to fuse the host membrane by irreversible conformational changes. The second cleavage site, S2′, located 130 residues from the N terminus of the S2 subunit which is highly conserved among coronaviruses. Cleavage at the S2′ site by host cell proteases is important for successful viral infection (Belouzard et al., 2009; Gui et al., 2017; Park et al., 2016b; Walls et al., 2017). The RBD is a core that mediates the interaction between S protein and ACE2 (Lan et al., 2020; Sternberg and Naujokat, 2020). Specifically, the S protein N-terminal S1 subunit mediates ACE2 binding whereas the C-terminal S2 subunit facilitates membrane fusion (Huang et al., 2020; Wrapp et al., 2020) to permit the transfer of the viral nucleocapsid into the target host cell (Belouzard et al., 2012; Lan et al., 2020). Recent computer modeling and structural analysis of the interaction between the SARS-CoV-2 RBD and ACE2 recognized the presence of residues important for ACE2 binding. Most of these residues are highly conserved or share similar side chain traits with those in the SARS-CoV RBD. However, those residues that mediate the SARS-CoV-2 RBD and ACE2 are experimentally unclear (Lan et al., 2020; Wan et al., 2020).

On account of its reported immunogenicity and solvent-exposed expression on the surface of the virus, the S protein has become a dominant target of various immune-based interventional and prophylactic strategies (Lan et al., 2020). Genetic analyses have played a significant role in expanding our knowledge on emerging viruses as well as informing viral containment strategies. With respect to the COVID19

pandemic, numerous, reoccurring mutations have been detected in the region coding for the S protein (van Dorp et al., 2020a). Functional analyses indicate that many of the mutations occurring in the S1 domain of the S protein alter virus transmissibility, infectivity, interaction with the target cells, and reactivity with neutralizing antibodies (Chatterjee, 2020; Focosi and Maggi, 2021; Greaney et al., 2021; Li et al., 2020). This genomic plasticity is related to the fact that viruses with RNA-based genomes are more prone to mutability compared to those with DNA-based genomes, and therefore evolve rapidly with selective pressures (Lin et al., 2019)., The genomic plasticity of RNA viruses sometimes enables the viral particle to elude neutralizing antibodies and virus-specific T cells generated post-infection or after vaccination (Focosi et al., 2021).

The antigenic heterogeneity caused by the high mutation rate display an unprecedented challenge in the production of successful vaccines as well as antisera and monoclonal antibody-based therapeutics (Servín-Blanco et al., 2016). Further attempts to confer immunity against viruses must therefore take ongoing antigenic variation into account, either through vaccine and immunotherapy solutions directed toward dominant viral genotypes, or inducing antibodies that identify a wide range of viral strains (Hedestam et al., 2008; Ledgerwood et al., 2015). The emergence of SARS-CoV-2 variants with ever-increasing mutations in the S protein will continue to challenge vaccine and immunotherapy solutions (McCormick et al., 2021). Thus, knowledge of dominant variants within the viral populations is essential for informing public health interventions. Research into the phylogeny and evolutionary process of the SARS-CoV-2 genotypes circulating in diverged geographical regions is critical in the initial stages of vaccine and immunotherapy design.

To investigate the impact of emerging variants on current vaccine and immunotherapies, we focused on the detection and distribution pattern of single nucleotide polymorphism (SNP) within the S protein from whole-genome sequence levels of SARS-CoV-2 strains from a closed population (Iran) with a narrow immunogenetic profile. Furthermore, evolutionary selection pressure on the viral population and the phylogenetic analyses of Iranian isolates of SARS-CoV-2 were generated to compare with other global isolates. Based on these investigations, we provide a clear image of the current dynamics of the COVID19 outbreak in Iran and evaluate the impact of emerging variants.

2. Materials and methods

2.1. Sequence selection and trimming

The RNA sequences of the whole genome and the S gene of SARS-CoV-2 genotypes were retrieved from the GISAID (Global Initiative on Sharing Avian Flu Data SOURCE). These data include 64 whole-genome sequences and 139 S gene sequences collected from different locations in Iran. In addition, 64 whole-genome sequences of various phylogenetic clades of other global isolates were retrieved from GISAID. SARS-CoV-2 sequences were trimmed to remove low-quality sequences with ambiguous nucleotides and to obtain sequences of the same size by ClustalW (version 2.1) implemented using Geneious Prime 2019 software (Biomatters, New Zealand). SNP identification, genetic diversity, and phylogenetic assessments were performed using whole-genome sequences and S coding sequences of SARS-CoV-2.

2.2. SNP discovery

SARS-CoV-2 sequences were mapped to the reference genome (NC_045512.2) for SNP identification and variant discovery via CLC Genomics Workbench (version 20, QIAGEN, Venlo, The Netherlands) followed by default parameters to the metrics of low-frequency variant detection; these metrics included SNP discovery Quality filter Neighborhood (radius 5), minimum central quality (20), minimum Neighborhood quality (15), minimum count (2) and minimum frequency (2%). SNPs were annotated and filtered using track tools and the Refseq

to understand their impacts on the amino acid change within the ORF2. Protein Data Bank (PDB) was downloaded from the RCSB PDB database (https://www.rcsb.org) for visualization of SNPs on the 3D structure of S protein and further prediction of their effect on drug binding sites in SARS-CoV-2 genome using CLC genomic workbench and RCSB PDB database.

The distribution pattern of SNPs in S gene sequences was assessed, and the evolutionary selection pressure based on computing the confidence estimation for the non-synonymous and synonymous nucleotide substitution rates (dN/dS = ω) and degree of selective constraints imposed on S protein were determined via the bootstrap method (1000 replicates) and Tamura-Nei model using MEGA version 7. The estimated Transition/Transversion bias was calculated for whole-genome and S gene data under the Kimura 2-parameter model using the maximum likelihood method. To the estimated value of the shape parameter for the discrete Gamma Distribution for whole-genome and S gene data Tamura-Nei model and Maximum Likelihood method were used in MEGA v.7.

2.3. Phylogenetic assessments

A phylogenetic tree based on the ClustalW alignment of all retrieved SARS-CoV-2 consensus sequences was constructed using MEGA v.7 (Pennsylvania State University, USA). The software default was set for the neighbor-joining method; maximum composite likelihood-parameter distance matrix, bootstrap values of 1000 replicates, and a 70% threshold score (Kumar et al., 2016).

2.4. Evaluation of Physicochemical parameters

Different Physico-chemical features of newly emerged viral variants including B.1.1.7 (Alpha), B.1.525 (Eta), and B.1.617 (Delta) were characterized and compared with Wuhan isolate (NC_045512) through subjecting S protein sequence of the viral variant to ProtParam server (https://web.expasy.org/protparam/) available at the ExPASy (Expert Protein Analysis System) bioinformatics resource portal. Various protein parameters including the molecular weight of the peptide, theoretical pI, instability index, grand average of hydropathicity, and total number of negatively/positively charged residues were estimated (https://web.expasy.org/protparam/).

3. Results and discussion

3.1. The pattern of SNPs in S the gene

Due to the rampant transmission of the virus throughout the human population, SARS-CoV-2 has the potential to be affected by high rates of recombination that might lead to new virulent derivatives of the virus. However, to date, there have been no recombination events in SARS COV-2 reported (Dearlove et al., 2020; Rausch et al., 2020). Genetic alteration in the genes encoding SPs such as the S protein could potentially aid the virus in evasion of the host immune response and diminish vaccine efficacy (Korber et al., 2020b). To assess this, we focused on the S protein of 189 SARS-CoV-2 isolates (genomic and sub-genomic) from Iran annotated based on the sequences recorded in the GISAID database (Supplementary Table 1). Initially, we observed that the frequency of the SNPs increased over time but then decreased to a rate similar to that of early isolates (Supplementary Table 2). Some SNPs, such as those positioned within nucleotide 614, were repeated in the S protein of most analyzed genotypes, indicating fixation and the emergence of prioritized SNPs at the S gene-level for natural selection. The SNPs of S protein analyzed for Iranian SARS-CoV-2 genotypes were mostly positioned in the NTD and were less frequent in the RBD domain. SNPs occurring within the S gene could potentially affect protein structure, antigenicity, and host tropism. Several mutations have been detected in the NTD of S protein of SARS-CoV-2 genotypes from different countries/regions. For

instance, Yuan et al. (2020) showed that five SNPs were located in residues of RBD, among which V483A (n = 21) in the USA and N439K (n = 31) in the UK were in RBM and 3 SNPs including A344S (n = 2) in Saudi Arabia, N354D (n = 2) in China and V367F (n = 8) in France and the Netherlands were in RBD (Yuan et al., 2020).

SNP variation in SARS-CoV-2 genome and their effect on the vaccine and anti-viral therapeutics.

SNP profiling revealed 112 SNPs in the S protein for Iranian isolate that led to an amino acid substitution (Supplementary Table 2). We determined a threshold of two repetitions, thus some SNPs were observed with low frequency. After that, we selected 36 important SNPs that occur in locations that putatively influence vaccines, antibody therapy and drugs (Table 1). Most amino acid-changing polymorphisms were positioned at the NTD coding region, while seventeen amino acid-changing polymorphisms were observed at the RBD of the S sequence (Table 1). The location of SNPs on the 3D structure of S protein is indicated in green (Fig. 1).

Most of the 25 SNPs were located within the S1 domain, while just 8 important SNPs were recognized in the NTD region of S1. The NTD has a role in the prefusion-to-postfusion transition of the virion (Chi et al., 2020). Besides, in many coronaviruses, the NTD domain of S protein attaches to host sialic acid receptors, and variations in NTD of coronaviruses have been shown to influence viral pathogenicity (Jaimes et al., 2020b; Millet et al., 2021). Nine SNPs in the Iranian isolates were located in the RBM region of the S sequence encoding the ACE2 receptor-binding domain (Table 1). Wang et al. (2021) have previously reported that more than half of all mutations on the RBD occurred in the RBM domain. These mutations may potentially strengthen the binding of S protein and ACE2 and impact antiviral drug and vaccine development, thus leading to more deleterious SARS-CoV-2 genotypes (Chen et al., 2020a; Wang et al., 2021).

The most prevalent SNP in Iranian isolates is D614G. This mutation has been also reported for the European strains of SARS-CoV-2 in the early stages of the pandemic and has become a principal substitution globally (Grubaugh et al., 2020). It has been speculated that the relative load of the G614 variant is higher than the parental D614 during infection of the upper respiratory tract. Therefore, this substitution seems to enhance the infectivity and transmissibility of the virus with no significant effect on disease severity (Korber et al., 2020b).

Several SNPs occurring in the S protein with lower frequency were also identified (Supplementary Table 2). Low-frequency SNPs can potentiate variations in the viral genome and viral tropism; however, understanding their effect on S protein function requires further investigation (Millet and Whittaker, 2015; Shang et al., 2020a; Shang et al., 2020b). Some S protein SNPs that we identified in this study were located within immunodominant epitopes that were previously suggested as targets for vaccine development (Ghorbani et al., 2020b) (Supplementary Table 2). However, the impact of these SNPs on humoral immunity and vaccine efficacy cannot be concluded without further clinical analysis.

Most of the SNPs share a constellation of mutations that is very similar to that occurring in the widely circulating variants B.1.1.7 (Alpha), B.1.525 (Eta) and B.1.617 (Delta). The accumulation of multiple mutations at these loci at the same time within a variant, especially in the area of interaction with the ACE2, could impact vaccines and immunotherapy-based treatment as well as naturally acquired immunity. Variant B.1.1.7 was first detected in England (2020-09-09) and quickly became the most prevalent variant in the UK, soon spreading to other countries. This variant possesses 17, then novel, genetic changes and a higher reproduction rate than previously described in other variants (Iacobucci, 2021).

Several mutations of SNPs occurring in non-S ORFs could be surface-exposed, indicating the potential capacity to interfere with antiviral drugs and therapies. SNP variations of SARS-CoV-2 whole genome of Iranian isolates with amino acid changes were found in ORFs 1ab (30 SNPs), 3a (4 SNPs), M (one SNP), 7b (2 SNPs), and N (9 SNPs) (Table 2).

Table 1
Important single-nucleotide polymorphisms (SNPs) in spike sequences of Iranian human SARS-COV-2 isolates and their effect on drug binding site.

Spike domain	Reference	Allele	Count	Frequency	Amino acid change	Other variants within codon	Effect on drug binding site	Modified positions are the same as the variants	The mutation that influences antibody responses
NTD	C	G	6	3.174603	Thr19Arg	No	1 drug hit	B.1.617	+
NTD	G	A	2	1.058201	Arg21Lys	No	No drug hits	B.1.617	+
NTD	TACATG	–	11	5.820106	His69_Val70del	Yes	No drug hits	B.1.1.7, B.1.525	+
NTD	G	T	4	2.116402	Asp80Tyr	No	No drug hits	B.1.351	+
NTD	TTA	–	20	10.58201	Tyr145del	No	1 drug hit	B.1.1.7, B.1.525	+
NTD	G	C	2	1.058201	Glu156Asp	Yes	No drug hits	B.1.617	+
NTD	AG	CT	2	1.058201	Arg158Leu	Yes	No drug hits	B.1.617	+
NTD	C	T	4	2.116402	Ala222Val	No	No drug hits	B.1.617	+
RBD	G	C	2	1.058201	Val407Leu	No	1 drug hit	–	+
RBD	G	A	2	1.058201	Ala411Thr	No	No drug hits	–	+
RBD	A	C	3	1.587302	Lys417Thr	Yes	1 drug hit	B.1.351, B.1.1.28	+
RBD (RBM)	C	G	3	1.587302	Asn439Lys	No	1 drug hit	–	+
RBD (RBM)	T	A	2	1.058201	Asn448Lys	No	No drug hits	–	+
RBD (RBM)	T	G	6	3.174603	Leu452Arg	No	1 drug hit	B.1.617, B.1.429 + B.1.427	+
RBD (RBM)	G	A	2	1.058201	Asp467Asn	No	1 drug hit	–	+
RBD (RBM)	G	A	15	7.936508	Ser477Asn	No	No drug hits	B.1.526	+
RBD (RBM)	C	A	6	3.174603	Thr478Lys	No	No drug hits	B.1.617	+
RBD (RBM)	G	A	3	1.587302	Glu484Lys	No	No drug hits	B.1.351, B.1.1.28, B.1.525	+
RBD (RBM)	A	T	23	12.16931	Asn501Tyr	Yes	1 drug hit	B.1.1.7, B.1.351, P.1	+
RBD (RBM)	A	C	2	1.058201	Asn501Thr	Yes	1 drug hit	B.1.351, B.1.1.28	+
RBD	G	A	2	1.058201	Arg509Lys	No	11 drug hits	–	+
RBD	C	A	2	1.058201	His519Asn	No	2 drug hits	–	+
RBD	G	C	2	1.058201	Ala522Pro	No	3 drug hits	–	+
RBD	T	G	2	1.058201	Val534Gly	Yes	2 drug hits	–	+
RBD	AA	GT	2	1.058201	Lys535Val	Yes	No drug hits	–	+
	C	A	16	8.465608	Ala570Asp	No	1 drug hit	B.1.1.7	+
	A	G	95	50.26455	Asp614Gly	No	1 drug hit	B.1.17, B.1.351, P.1, B.1.525, B.1.429 + B.1.427, B.1.526, B.1.617, B.1.1.28	+
	C	T	4	2.116402	His655Tyr	No	1 drug hit	B.1.1.28	+
	G	T	3	1.587302	Gln677His	No	No drug hits	B.1.525	+
	C	A	16	8.465608	Pro681His	No	1 drug hit	B.1.1.7	+
	C	G	6	3.174603	Pro681Arg	No	1 drug hit	B.1.617	+
	C	T	16	8.465608	Thr716Ile	No	1 drug hit	B.1.1.7	+
HR	G	A	8	4.232804	Asp950Asn	No	No drug hits	B.1.617	+
HR	T	G	17	8.994709	Ser982Ala	No	No drug hits	B.1.1.7	+
	G	C	16	8.465608	Asp1118His	No	No drug hits	B.1.1.7	+

NTD; N-terminal domain (NTD), RBD; receptor-binding domain, RBM; receptor-binding motif, FP; fusion peptide, HR; heptad repeat1.

Saha et al. (2021) also reported SNPs in ORFs 1ab, 3a, M, and N genes of 566 genotypes of SARS-CoV-2 from India with the potential of amino acid alteration (Saha et al., 2021). When the number of SNPs was normalized according to ORF length, 7b, S, and N proteins showed more variability compared to other ORFs. N is an important protein in the disease cycle and based on viral gene expression analysis, exhibits a higher level of expression during cell infection (Ghorbani et al., 2020b).

The study of mutations within each country can provide valuable information for the development of vaccines and immunotherapies and new insight into disease development within a given country. This is especially valuable to assess in countries consisting of closed populations, as the effects of more homogeneous immunogenetic properties of the population can be more carefully assessed. In this study, the SNP variation that has been detected in the S and 1ab proteins can affect the drug binding site. *N*-acetylglucosamine (NAG), polysorbate 80, Isoleucine, Glycine, and Lysine were affected by these mutations. NAG and Glycine are under investigation for therapeutic potential, while Lysine has been used prophylactically (Quantinosis.aiLLC (2021); Vargas, 2020). Polysorbate 80 is a non-ionic surfactant and emulsifier often used in foods and cosmetics and is a component of many vaccines used in the

United States, including the Janssen COVID-19 vaccine. Fig. 2 displays the interaction of *N*-acetylglucosamine (NAG) and polysorbate 80 with S protein in positions that were under SNP variation. NAG has been affected by SNPs more than other drugs.

The SNPs in other proteins of SARS-CoV-2 genotypes recorded from Iran were determined and drug binding sites were investigated (Table 2). NAG is used in the management of numerous disease states including osteoarthritis, diabetes, aging skin, knee pain, and inflammatory bowel disease (IBD) and phase one clinical assessment of its use in the therapeutic intervention of COVID-19 is currently under investigation (https://clinicaltrials.gov/ct2/show/NCT04706416).

Neutralizing monoclonal antibodies and targeting the RBD on the SARS-CoV-2 S protein are potential options for drug development for treating COVID-19. Therefore, monoclonal antibodies targeting S1 block the viral entry to host cells (Chen et al., 2020b). In this study, we showed the mutation that can affect the S protein antibody binding sites (Table 1). The Seventeen SNP have occurred on the RBD of Iranian isolates which can affect the targeting of the RBD. The mutations K417N/T, N439K, L452R, S477N, E484K, and N501Y were reported to be the most dangerous for immune escape from antibody blocking
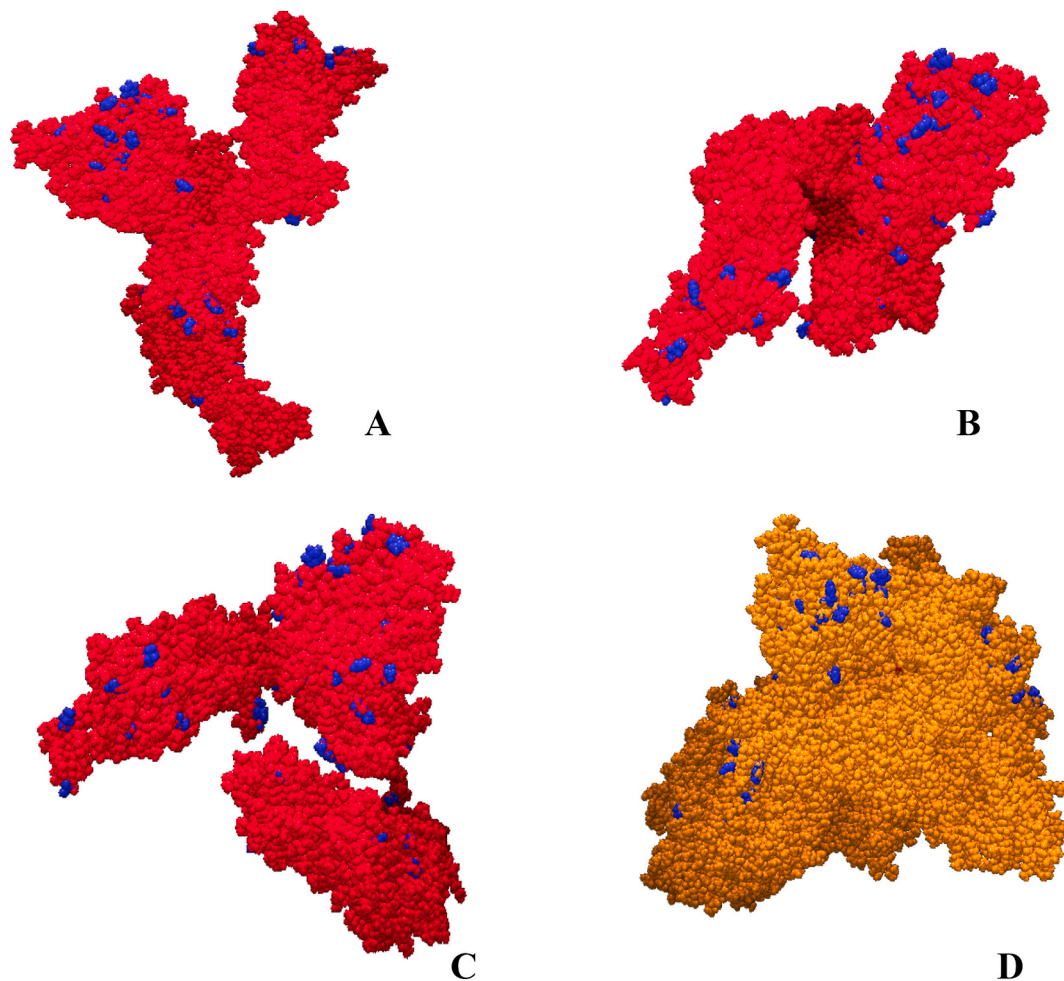
Fig. 1. Position of SNPs on the 3D structure of SARS-CoV-2 Spike protein in Iranian isolates. SNPs have been shown in green color. A, B, and C show SNPs in a chain. D show SNPs in a chain that combined to other two chains. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Focosi and Maggi, 2021), mutations were discovered in our analysis (Table 1). Accumulation SNPs in RBD can affect antibody therapeutics and escape antibody binding (Greaney et al., 2021).

3.2. Phylogeny and evolution analysis of SARS-CoV-2 genome

Addressing specific types of mutations: for viruses in general, nucleotide substitutions are on average four times more common than insertions/deletions (Sanjuán, 2010). For SARS-CoV-2, the mutation rate is estimated at 8×10^{-4} to 1.1×10^{-3} substitutions/site/ (Duchene et al., 2020); This corresponds to an average pairwise nucleotide difference across any isolates of 8–10 (van Dorp et al., 2020a). The transition to transversion ratio for point mutations is considered a good indicator of the evolutionary pressure on a given virus, and for SARS-CoV-2, the genome-wide ratio is calculated at 1.88 (van Dorp et al., 2020b). However, the transition/transversion ratio may vary in different genes within a viral population (Strandberg and Salter, 2004).

Our results revealed that the transition/transversion ratio bias for the S encoding gene is lower than the whole genome of SARS-CoV-2 (Table 3), indicating that the evolutionary pressure is focused on conserving the S protein. However, a high transition rate, in hotspot regions, could lead to a positive selection of S mutations associated with virulence properties, resistance against host immunity, and infectivity of the virus. Roy et al. (2020) reported that the frequency of transition changes in SARS-CoV-2 was higher than transversion in the pan-genome of the virus. They concluded that mutations related to non-structural

protein-coding genes of SARS-CoV-2 are under negative selection, while mutations related to structural protein-coding genes are under positive selection (Roy et al., 2020). Gamma parameter for site rates was calculated and our data showed that gamma parameter for S protein is higher than the whole-viral genome; thus more positive selective pressure is on S protein and this may be related to selection pressure exerted by the adaptive immune response (Table 3) (Gelman et al., 2020). SNPs of SARS-CoV-2 naturally exist in the population (Ghorbani et al., 2020a) or accumulate in a new variant when the virus circulates in different hosts (Ghorbani et al., 2021)but their frequency is related to positive selection by mAbs and vaccines (Gelman et al., 2020).

A phylogenetic tree was constructed for selected SARS-CoV-2 isolates from Iran and compared with different clades of SARS-CoV-2 genotypes from other countries and isolates of new variants that were constructed based on the whole-genome sequences of the virus by the NJ method. Most Iranian isolates showed a close evolutionary relationship to other viral genotypes from Iran (Fig. 3).

For phylogenetic analysis, we assessed the genomic diversity of the Iranian SARS-CoV-2 isolates and their phylogenetic relationship with other strains from various parts of the globe. Based on the evolutionary relationship shown in the phylogenetic NJ tree for the isolates in this study, most of the isolates in Iran were clustered in close clades and rather distally from the out-group (SARS-related coronavirus), which could be due to geographical separation among countries and internal circulation and adaptation of the virus in Iran. A small number of isolates were clustered in other clades related to other regions over the

Table 2
Single-nucleotide polymorphisms (SNPs) in whole genome (except spike ORF) sequences of Iranian human SARS-COV-2 isolates and their effect on drug binding site.

ORF	Reference	Allele	Count	Frequency	Amino acid change	Effect on putative drug-binding site (PDB drug name)
ORF1ab	G	T	2	3.125	Trp161Leu	–
ORF1ab	C	T	16	25	Arg207Cys	–
ORF1ab	G	C	2	3.125	Cys316Ser	–
ORF1ab	G	A	36	56.25	Val378Ile	–
ORF1ab	C	T	6	9.375	His417Tyr	–
ORF1ab	A	C	2	3.125	Glu489Ala	–
ORF1ab	C	T	2	3.125	Ala656Val	–
ORF1ab	C	T	3	4.6875	Ser944Leu	–
ORF1ab	T	C	2	3.125	Cys1114Arg	–
ORF1ab	C	T	3	4.6875	Leu1507Phe	–
ORF1ab	C	T	2	3.125	Thr1682Ile	–
ORF1ab	C	T	3	4.6875	Thr1760Ile	–
ORF1ab	G	T	2	3.125	Arg2159Leu	–
ORF1ab	C	T	2	3.125	Ser2488Phe	–
ORF1ab	A	G	2	3.125	Asn2596Ser	–
ORF1ab	G	T	16	25	Met2796Ile	–
ORF1ab	G	T	32	50	Leu3606Phe	–
ORF1ab	C	A	2	3.125	His3633Asn	–
ORF1ab	A	G	3	4.6875	Glu3909Gly	–
ORF1ab	C	T	2	3.125	Thr4129Ile	–
ORF1ab	G	T	3	4.6875	Gly4529Cys	–
ORF1ab	C	T	27	42.1875	Pro4715Leu	–
ORF1ab	C	T	2	3.125	Leu5030Phe	–
ORF1ab	C	T	6	9.375	Thr5193Ile	–
ORF1ab	C	T	4	6.25	His5614Tyr	UR7, VW4, UXG, N0E, S9S, UQS, VVD, JFM, VWD, NZG, VWJ, VWG, K34*
ORF1ab	C	T	3	4.6875	Thr5675Ile	VVG, STV
ORF1ab	T	G	2	3.125	Phe5799Val	K34
ORF1ab	C	T	2	3.125	Ser5809Leu	VWM, VXG, K2P
ORF1ab	C	T	2	3.125	Thr6038Ile	–
ORF1ab	G	A	6	9.375	Gly6875Arg	–
ORF3a	G	T	15	23.4375	Gln57His	–
ORF3a	G	T	3	4.6875	Val112Phe	–
ORF3a	A	G	3	4.6875	Thr223Ala	–
ORF3a	C	T	2	3.125	Thr223Ile	–
ORFM	A	G	3	4.6875	Ile73Met	–
ORF7b	G	T	3	4.6875	Leu11Phe	–
ORF7b	C	T	3	4.6875	Thr40Ile	–
ORFN	G	T	2	3.125	Asp3Tyr	–
ORFN	C	A	3	4.6875	Pro13Thr	–
ORFN	C	T	2	3.125	Ala35Val	–
ORFN	C	T	6	9.375	Ser186Phe	–
ORFN	C	G	2	3.125	Asn192Lys	–
ORFN	C	T	13	20.3125	Ser194Leu	–
ORFN	GGG	AAC	7	10.9375	Gly204delinsLysArg	–
ORFN	C	T	3	4.6875	Ala220Val	–
ORFN	G	T	2	3.125	Met234Ile	–

* UR7: 1-(3-fluoro-4-methylphenyl)methanesulfonamide, VW4: (2S)-2-phenylpropane-1-sulfonamide, UXG: 1-(diphenylmethyl)azetidin-3-ol, N0E: N}-(4-hydrox-yphenyl)-3-phenyl-propanamide, S9S: -[2-(4-fluorophenyl)ethyl]methanesulfonamide, UQS: N-[(2-fluorophenyl)methyl]-1H-pyrazol-4-amine, VVD: 5-(acetylamino)-2-fluorobenzoic acid, JFM: N-(2-phenylethyl) methanesulfonamide, VWD: (1R)-2-(methylsulfonyl)-1-phenylethan-1-ol, NZG: 3-(acetylamino)-4-fluorobenzoic acid, VWJ: N-(propan-2-yl)-1H-benzimidazol-2-amine, VWG: N-hydroxyquinoline-2-carboxamide, K34: 5-(1,3-thiazol-2-yl)-1H-1,2,4-triazole, VVG: N-(2-fluorophenyl) ethane sulfonamide, STV: -(1,3-benzodioxol-5-ylmethyl) ethane sulfonamide, VWM: (3R)-1-acetyl-3-hydroxypiperidine-3-carboxylic acid, VXG: (3S,4R)-1-acetyl-4-phenylpyrrolidine-3-carboxylic acid, K2P: 2-(trifluoromethoxy)benzoic acid.

world including England, Wales, and Yunnan, which could be due to the rapid and extensive transmission of the virus from one continent to another or parallel mutations in the viral populations. B.1.1.7, B.1.617 and B.1.525 variants have a close evolutionary relationship with Iranian isolates. This is the first reported confirmation of the presence of the B.1.1.7, B.1.617 and B.1.525 variants in Iran.

3.3. Protein physicochemical parameters

The knowledge of physicochemical properties of new SARS-CoV-2 variants particularly at the S protein level is vital for developing the live attenuated and inactivated vaccines against SARS-CoV-2 and to properly determining the drug-targeting strategies for small-molecule pharmaceuticals. Here, we have calculated important physicochemical parameters of the S protein of new SARS-CoV-2 variants including B.1.1.7 (Alpha), B.1.525 (Eta) and B.1.617 (Delta) and compared the calculated values with those of the Refseq genotype that was reported

from Wuhan at the beginning of COVID-19 pandemic (Table 4). The values for the physicochemical properties and the molecular weight of the (S) proteins were calculated based on the corresponding amino acid sequence. The physicochemical properties of the newly emerged variant of SARS-CoV-2 slightly differed from the original genotype because the main immunogenic properties of viral variants and the virus from Wuhan were not changed throughout ongoing evolution (Table 4). The grand average of the hydropathicity index for all variants shows that the S proteins of different variants hardly differ in their hydrophobicity. Since the response of viruses to disinfectants depends on whether they are lipophilic or hydrophilic, viruses can be categorized as lipophilic (enveloped) or hydrophilic (nonenveloped) and intermediate solubility (nonenveloped) (Block, 2001). SARS-CoV-2 and other coronaviruses have an envelope and are classified as lipophilic viruses (Koch, 1985). Srivastava et al. (2020) also found that the more lipophilic the drug, the better it can inhibit the SARS-CoV-2 replication within the infected human cells. In enveloped viruses, the viral protein and lipid
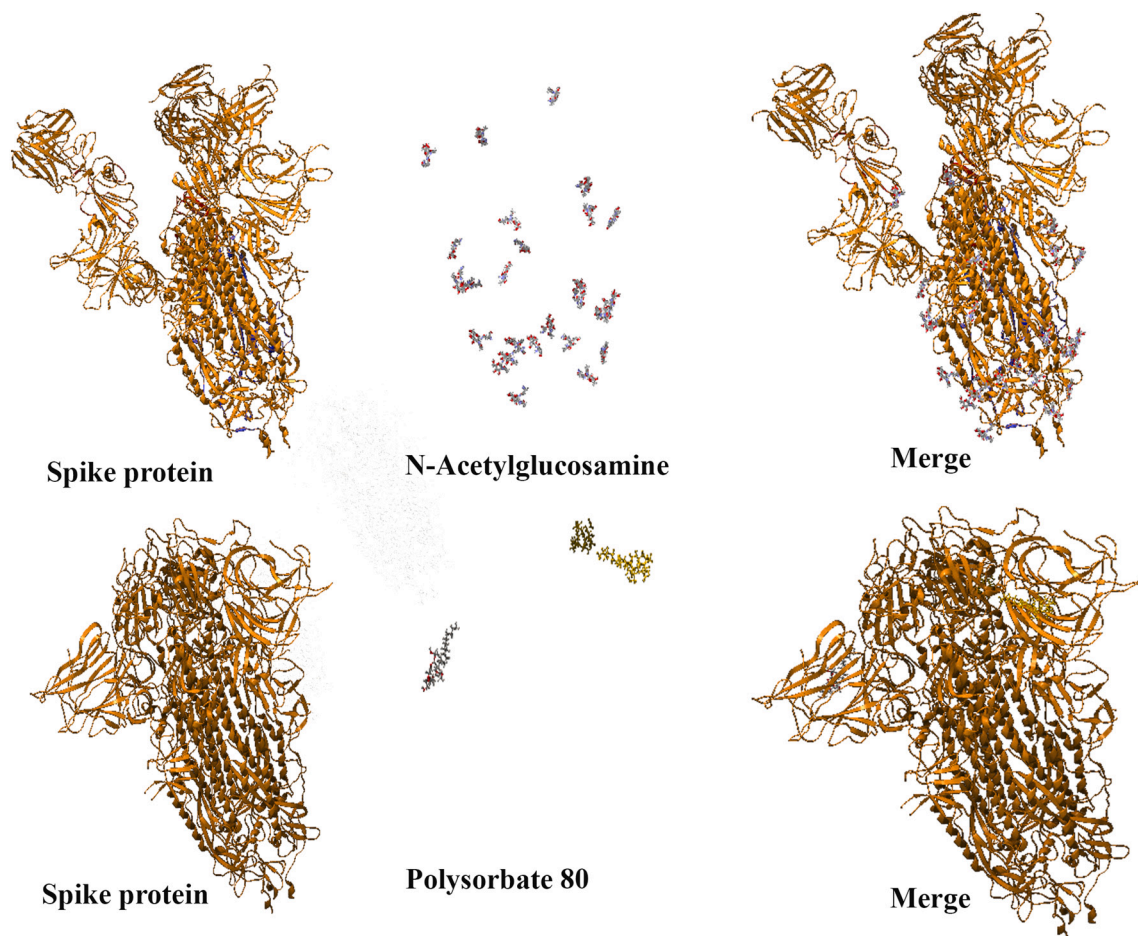
Fig. 2. Location of SNPs in the binding site of *N*-acetylglucosamine and polysorbate 80 with spike protein of SARS-CoV-2, Iranian isolates.

Table 3
Transition/transversion and Gamma parameter based on whole-genome and spike gene sequences of Iranian isolates.

Based on	Transition/transversion (R)	Gamma parameter for site rates
Whole genome	1.33	0.05
Spike gene	0.64	0.339

compositions, and the host cell membrane plays a decisive role in infectivity (Srivastava et al., 2020; Sun and Whittaker, 2003). The S2 subunit is composed of FP, HR1, HR2, TM domain, and cytoplasmic domain fusion (CT) responsible for viral fusion and entry. FP includes 15–20 conserved amino acids of the *coronaviridae* family and mainly of hydrophobic residues, such as glycine (G) or alanine (A), which anchor to the target membrane when the S protein adopts the prehairpin conformation. FP plays an essential role in mediating membrane fusion by disrupting and connecting lipid bilayers of the host cell membrane and possible active substances against B.1.617 (Delta) variant should therefore be of a further lipophilic nature in order to penetrate the membrane of this specific genotype and inactivate the virus (Millet and Whittaker, 2018; Srivastava et al., 2020). However, the efficiency of a drug to penetrate the viral membrane is not always directly related to the loss of the replication functionality of the nucleic acid and its complete demolition (Block, 2001).

Although the total structural charge of SARS-CoV-2 is positive, the SPs of SARS-CoV-2 are carrying varied total electric charges based on their amino acid content. The E, M and N proteins are positive, and the surface spike protein S is negative. This is consistent with our findings for other variants of SARS-CoV-2 as shown in Table 4 (Pawłowski,

2021). The instability index values of the S protein for all variants ranged from 32.58 to 33.01 which classifies S protein as stable within analyzed genotypes. The pI) for different viral variants ranged from 6.24 to 6.78 which was within the range discovered for the immunogenic epitopes of SARS-CoV-2 S protein reported (Li et al., 2021).

4. Conclusion

Superspreading events in which many people are infected at once, typically by a single individual, have shown to contribute to the rapid transmission of SARS-CoV-2. The more frequent and transmissible variants from the United Kingdom, South Africa and Brazil have pushed out other strains of SARS-CoV-2. Early introduction of new variants may lead to limited onward transmission. Even though the super spreading events can be devastating to the residents, they have limited large-scale impacts worldwide because they occur later and in a more isolated population (Lemieux et al., 2021). Since the long-term travel restrictions and border closures are not desirable, reducing the risk of introducing variants, and ensuring that those that are introduced do not effect the vaccine efficiency, will help countries to maintain low levels of SARS-CoV-2 transmission. Profiling SNPs and constant monitoring of selective pressure on viral population, introduction of new variants and understanding the factors that contribute to superspreading are crucial for maintaining the vaccine efficacy, preventing the breakthrough infections and new insights toward vaccines and anti-viral therapeutics (Lemieux et al., 2021; Lewis, 2021). Genomic analysis of the SARS-CoV-2 genotypes in Iran emphasizes the importance of superspreading events in shaping the course of this pandemic. Those residues of RBD of S protein that mostly affect the SARS-CoV-2 cell entry, are hotspots for
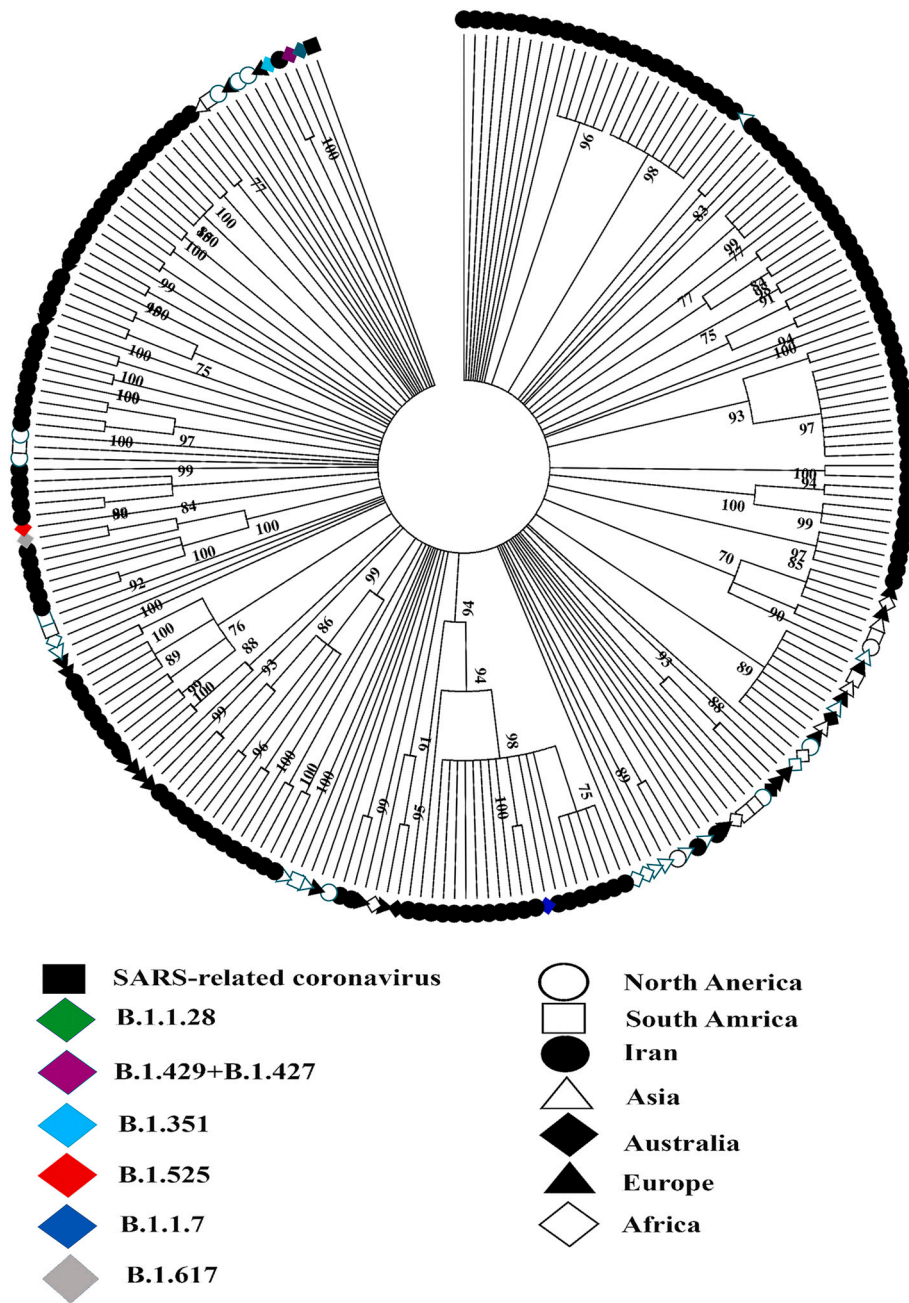
Fig. 3. The whole genome-based phylogenetic analysis of SARS-CoV-2 isolates from Iran and other countries and a candidate of new variants.

Table 4
Physicochemical parameters of major new SARS-CoV-2 variants B.1.1.7 (Alpha), B.1.525 (Eta) and B.1.617 (Delta).

Parameters	Molecular weight (kDa)	Theoretical pI	The instability index	Grand average of hydropathicity	Total number of negatively/positively charged residues
Wuhan	141.178	6.24	33.01	−0.079	110/103
B.1.1.7 (Alpha)	140.824	6.35	32.58	−0.075	109/103
B.1.525 (Eta)	141.150	6.32	32.86	−0.077	109/103
B.1.617 (Delta)	140.986	6.78	32.81	−0.090	108/106

mutation. Detected SNPs positioned within S1 and S2 domains may affect the fusion and pathogenicity of the virus. We observed mutations in the S1 gene, similar to new variants such as B.1.1.7, B.1.617 and B.1.525. The analyses conducted in this study revealed that the detected SNPs may affect drug and antibody binding sites within the whole genome of the virus. Not all case clusters were the result of super-spreading events, as some Iranian genotypes phylogenetically

belonged to clusters of unrelated genotype cases from neighboring countries. Similarities in SNP profiles seen within B.1.1.7, B.1.617 and B.1.525 that recently sequenced genotypes of SARS-CoV-2 from Iran raises the possibility that the superspreading in Iran may encompass varied transmission dynamics. It also suggests the role of chance in the trajectory of an epidemic. Our research indicates that a single intro-duction of a new genotype of the virus had a huge effect on subsequent

transmission as it was amplified through superspreading in a highly ambulant population early within the outbreak, before public health precautions limiting exponential growth and subsequent superspreading events.

Rapid changes in the SARS-CoV-2 population in Iran suggest that the likelihood of the appearance of new variants in this country is imminent unless the viral spread is controlled through rapid vaccination or social distancing measures. For example, the superspreading of D614G was of urgent concern; it began spreading in Europe in early February, and when introduced to new regions, quickly became the dominant strain. Whilst, for MERS-CoV, superspreading events were not associated with mutations in the virus sequences that drive increased transmission (Park et al., 2016a).

Moreover, evidence of recombination between strains indicates multiple strain infections which are important implications for SARS-CoV-2 transmission, pathogenesis and immune interventions (Korber et al., 2020a). For instance, Chinese researchers found that SARS-CoV-2 could be classified into two major local variants named L-type and S-type. L-type prevailed at the early stages of the outbreak in Wuhan, whilst the S-type was phylogenetically older than L-type and less prevalent at an early stage, but with a later increase in frequency in Wuhan (Awadasseid et al., 2021). Further insight into local variations, detection of variants from superspreadings, and their characteristics will benefit assessing risks and developing better treatment and prevention strategies.

Therefore, constant monitoring of genome mutations specifically of local stains is essential to understand the evolution of the SARS-CoV-2 genome under selection pressure. However, further experimental investigations are required to define the impact of the detected SNPs on pathogenicity and transmission of the virus for developing the appropriate controlling protocol, therapeutic, and vaccination strategies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genrep.2022.101537.

CRedit authorship contribution statement

A.G., S·S and LB conceived and designed the experiments, A.G. analyzed the data, A.G., S·S., LB., TPK., MJ and MB wrote the paper; AA., MHS., AN., TPK, S.S and KI the edited paper that was approved by all authors.

Declaration of competing interest

The authors declare that they have no conflict of interest. The research reported here did not involve experimentation with human participants or animal.

References

Awadasseid, A., Wu, Y., Tanaka, Y., Zhang, W., 2021. SARS-CoV-2 variants evolved during the early stage of the pandemic and effects of mutations on adaptation in Wuhan populations. Int. J. Biol. Sci. 17, 97.

Belouzard, S., Chu, V.C., Whittaker, G.R., 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl. Acad. Sci. 106, 5871–5876.

Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4, 1011–1033.

Block, S.S., 2001. Disinfection, Sterilization, And Preservation. Lippincott Williams & Wilkins.

Chatterjee, S., 2020. An overview of mutations occurring within the coronavirus-2 genome: mutations data reporting on SARS-CoV-2. Available at SSRN 3632241.

Chen, J., Wang, R., Wang, M., Wei, G.-W., 2020a. Mutations strengthened SARS-CoV-2 infectivity. J. Mol. Biol. 432, 5212–5226.

Chen, X., Li, R., Pan, Z., Qian, C., Yang, Y., You, R., Zhao, J., Liu, P., Gao, L., Li, Z., 2020b. Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. Cell.Mol.Immunol. 17, 647–649.

Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y., 2020. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science 369, 650–655.

Dearlove, B., Lewitus, E., Bai, H., Li, Y., Reeves, D.B., Joyce, M.G., Scott, P.T., Amare, M. F., Vasan, S., Michael, N.L., 2020. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proc. Natl. Acad. Sci. 117, 23652–23662.

van Dorp, L., Acman, M., Richard, D., Shaw, L.P., Ford, C.E., Ormond, L., Owen, C.J., Pang, J., Tan, C.C., Boshier, F.A., 2020a. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect. Genet. Evol. 83, 104351.

van Dorp, L., Richard, D., Tan, C.C., Shaw, L.P., Acman, M., Balloux, F., 2020b. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat. Commun. 11, 1–8.

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., Baele, G., 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. VirusEvol. 6, veaa061.

Feng, W., Zong, W., Wang, F., Ju, S., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): a review. Mol. Cancer 19, 1–14.

Focosi, D., Maggi, F., 2021. Neutralising antibody escape of SARS-CoV-2 spike protein: risk assessment for antibody-based Covid-19 therapeutics and vaccines. Rev. Med. Virol. 31, e2231. https://doi.org/10.1002/rmv.2231.

Focosi, D., Maggi, F., Mazzetti, P., Pistello, M., 2021. Viral infection neutralization tests: a focus on severe acute respiratory syndrome-coronavirus-2 with implications for convalescent plasma therapy. Rev. Med. Virol. 31, e2170.

Gelman, R., Bayatra, A., Kessler, A., Schwartz, A., Ilan, Y., 2020. Targeting SARS-CoV-2 receptors as a means for reducing infectivity and improving antiviral and immune response: an algorithm-based method for overcoming resistance to antiviral agents. Emerg.MicrobesInfect. 9, 1397–1406.

Ghadiri, F., Sahraian, M.A., Saeedi, R., Moghadasi, A.N., 2021. Attitudes toward vaccination in patients with multiple sclerosis:a report from Iran. Mult. Scler. Relat. Disord. 53, 103045. https://doi.org/10.1016/j.msard.2021.103045.

Ghorbani, A., Samarfard, S., Ramezani, A., Izadpanah, K., Afsharifar, A., Eskandari, M. H., Karbanowicz, T.P., Peters, J.R., 2020a. Quasi-species nature and differential gene expression of severe acute respiratory syndrome coronavirus 2 and phylogenetic analysis of a novel Iranian strain. Infect. Genet. Evol. 85, 104556.

Ghorbani, A., Zare, F., Sazegari, S., Afsharifar, A., Eskandari, M., Pormohammad, A., 2020b. Development of a novel platform of virus-like particle (VLP)-based vaccine against COVID-19 by exposing epitopes: an immunoinformatics approach. New MicrobesNew Infect. 38, 100786.

Ghorbani, A., Samarfard, S., Eskandarzade, N., Afsharifar, A., Eskandari, M.H., Niazi, A., Izadpanah, K., Karbanowicz, T.P., 2021. Comparative phylogenetic analysis of SARS-CoV-2 spike protein—possibility effect on virus spillover. Brief. Bioinform. 1–9. https://doi.org/10.1093/bib/bbab144.

Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H., 2021. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host Microbe 29 (44–57), e49.

Greer, S.L., King, E.J., da Fonseca, E.M., Peralta-Santos, A., 2020. The comparative politics of COVID-19: the need to understand government responses. Glob.Public Health 15, 1413–1416.

Grubaugh, N.D., Hanage, W.P., Rasmussen, A.L., 2020. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. Cell 182, 794–795.

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., Wang, X., 2017. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res. 27, 119–129.

Hadianfar, A., Yousefi, R., Delavary, M., Fakoor, V., Shakeri, M.T., Lavallière, M., 2021. Effects of government policies and the Nowruz holidays on confirmed COVID-19 cases in Iran: an intervention time series analysis. PloS One 16, e0256516.

Harapan, H., Itoh, N., Yufika, A., Winardi, W., Keam, S., Te, H., Megawati, D., Hayati, Z., Wagner, A.L., Mudatsir, M., 2020. Coronavirus disease 2019 (COVID-19): a literature review. J. Infect. Public Health 13 (5), 667–673. https://doi.org/10.1016/j.jiph.2020.03.019.

Hedestam, G.B.K., Fouchier, R.A., Phogat, S., Burton, D.R., Sodroski, J., Wyatt, R.T., 2008. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. Nat. Rev. Microbiol. 6, 143–155.

Huang, Y., Yang, C., Xu, X.-F., Xu, W., Liu, S.-W., 2020. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol. Sin. 41, 1141–1149.

Iacobucci, G., 2021. Covid-19: new UK variant may be linked to increased death rate, early data indicate. BMJ 372 (230).

Jaimes, J.A., André, N.M., Chappie, J.S., Millet, J.K., Whittaker, G.R., 2020a. Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. J. Mol. Biol. 432, 3309–3325.

Jaimes, J.A., Millet, J.K., Whittaker, G.R., 2020b. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site. IScience 23, 101212.

Koch, F., 1985. Disinfection, Sterilization, And Preservation. Wiley Online Library.

Korber, B., Fischer, W., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E.E., Bhattacharya, T., Parker, M.D., 2020. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. BioRxiv.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., 2020b. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182 (812–827), e819.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220.

Ledgerwood, J., Coates, E., Yamshchikov, G., Saunders, J., Holman, L., Enama, M., DeZure, A., Lynch, R., Gordon, I., Plummer, S., 2015. Safety, pharmacokinetics and neutralization of the broadly neutralizing HIV-1 human monoclonal antibody VRC01 in healthy adults. Clin.Exp.Immunol. 182, 289–301.

Lemieux, J.E., Siddle, K.J., Shaw, B.M., Loreth, C., Schaffner, S.F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C.H., Krasilnikova, L.A., 2021. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science 371, eabe3261.

Lewis, D., 2021. Superspreading drives the COVID pandemic-and could help to tame it. Nature 544–546.

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., 2020. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 182 (1284–1294), e1289.

Li, Y., Ma, M.-L., Lei, Q., Wang, F., Hong, W., Lai, D.-Y., Hou, H., Xu, Z.-W., Zhang, B., Chen, H., 2021. Linear epitope landscape of the SARS-CoV-2 spike protein constructed from 1,051 COVID-19 patients. Cell Rep. 34, 108915.

Lin, J.-J., Bhattacharjee, M.J., Yu, C.-P., Tseng, Y.Y., Li, W.-H., 2019. Many human RNA viruses show extraordinarily stringent selective constraints on protein evolution. Proc. Natl. Acad. Sci. 116, 19009–19018.

McCormick, K.D., Jacobs, J.L., Mellors, J.W., 2021. The emerging plasticity of SARS-CoV-2. Science 371, 1306–1308.

Millet, J.K., Whittaker, G.R., 2015. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. Virus Res. 202, 120–134.

Millet, J.K., Whittaker, G.R., 2018. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. Virology 517, 3–8.

Millet, J.K., Jaimes, J.A., Whittaker, G.R., 2021. Molecular diversity of coronavirus host cell entry receptors. FEMS Microbiol. Rev. 45, fuaa057.

Park, D., Huh, H.J., Kim, Y.J., Son, D.-S., Jeon, H.-J., Im, E.-H., Kim, J.-W., Lee, N.Y., Kang, E.-S., Kang, C.I., 2016a. Analysis of intrapatient heterogeneity uncovers the microevolution of Middle East respiratory syndrome coronavirus. Mol.Case Stud. 2, a001214.

Park, J.-E., Li, K., Barlan, A., Fehr, A.R., Perlman, S., McCray, P.B., Gallagher, T., 2016b. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. Proc. Natl. Acad. Sci. 113, 12262–12267.

Pawłowski, P.H., 2021. Charged amino acids may promote coronavirus SARS-CoV-2 fusion with the host ce. AIMS Biophys. 8, 111–120.

Quantinosis.aiLLC, 2021. N-acetyl Glucosamine as Therapeutic Intervention for Coronavirus Disease-19 (COVID-19) ClinicalTrials.gov Identifier: NCT04706416. Quantinosis.ai LLC.

Rausch, J.W., Capoferri, A.A., Katusiime, M.G., Patro, S.C., Kearney, M.F., 2020. Low genetic diversity may be an Achilles heel of SARS-CoV-2. Proc. Natl. Acad. Sci. 117, 24614–24616.

Rella, S.A., Kulikova, Y.A., Dermitzakis, E.T., Kondrashov, F.A., 2021. Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains. Sci. Rep. 11, 1–10.

Roy, C., Mandal, S.M., Mondal, S.K., Mukherjee, S., Mapder, T., Ghosh, W., Chakraborty, R., 2020. Trends of mutation accumulation across global SARS-CoV-2 genomes: implications for the evolution of the novel coronavirus. Genomics 112, 5331–5342.

Saha, I., Ghosh, N., Pradhan, A., Sharma, N., Maity, D., Mitra, K., 2021. Whole genome analysis of more than 10 000 SARS-CoV-2 virus unveils global genetic diversity and target region of NSP6. Brief. Bioinform. 22, 1106–1121.

Sanjuán, R., 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos.Trans.R.Soc. B: Biol.Sci. 365, 1975–1982.

Servín-Blanco, R., Zamora-Alvarado, R., Gevorkian, G., Manoutcharian, K., 2016. Antigenic variability: obstacles on the road to vaccines against traditionally difficult targets. Hum.VaccinesImmunother. 12, 2640–2648.

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., Li, F., 2020a. Cell entry mechanisms of SARS-CoV-2. Proc. Natl. Acad. Sci. 117, 11727–11734.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., Li, F., 2020b. Structural basis of receptor recognition by SARS-CoV-2. Nature 581, 221–224.

Shaw, L.P., Wang, A.D., Dylus, D., Meier, M., Pogacnik, G., Dessimoz, C., Balloux, F., 2020. The phylogenetic range of bacterial and viral pathogens of vertebrates. Mol. Ecol. 29, 3361–3379.

Sola, I., Almazan, F., Zuniga, S., Enjuanes, L., 2015. Continuous and discontinuous RNA synthesis in coronaviruses. Ann.Rev.Virol. 2, 265–2883.

Srivastava, A.K., Kumar, A., Tiwari, G., Kumar, R., Misra, N., 2020. In silico investigations on the potential inhibitors for COVID-19 protease arXiv preprint arXiv:2003.10642.

Sternberg, A., Naujokat, C., 2020. Structural features of coronavirus SARS-CoV-2 spike protein: targets for vaccination. Life Sci. 118056.

Strandberg, A.K., Salter, L.A., 2004. A comparison of methods for estimating the transition: transversion ratio from DNA sequences. Mol. Phylogenet. Evol. 32, 495–503.

Sun, X., Whittaker, G.R., 2003. Role of influenza virus envelope cholesterol in virus entry and infection. J. Virol. 77, 12543–12551.

Vargas, M.H., 2020. Glycine Supplement for Severe COVID-19 (ClinicalTrials.gov Identifier: NCT04443673). Instituto Nacional de Enfermedades Respiratorias.

Walls, A.C., Tortorici, M.A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F.A., Veesler, D., 2017. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proc. Natl. Acad. Sci. 114, 11157–11162.

Wan, Y., Shang, J., Graham, R., Baric, R.S., Li, F., 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J. Virol. 94, e00127-00120.

Wang, R., Chen, J., Gao, K., Hozumi, Y., Yin, C., Wei, G.-W., 2021. Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. Commun.Biol. 4, 1–14.

Wong, L.P., Alias, H., Danaee, M., Ahmed, J., Lachyan, A., Cai, C.Z., Lin, Y., Hu, Z., Tan, S.Y., Lu, Y., 2021. COVID-19 vaccination intention and vaccine characteristics influencing vaccination acceptance: a global survey of 17 countries. Infect.Dis. Poverty 10, 1–14.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B. S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269.

Yuan, F., Wang, L., Fang, Y., Wang, L., 2020. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. Transbound. Emerg. Dis. 68 (6), 3288–3304. https://doi.org/10.1111/tbed.13931.