Study on SARS-CoV-2 strains in Iran reveals potential contribution of co-infection with and recombination between different strains to the emergence of new strains

Peyman Taghizadeh [a], Sadegh Salehi [b], Ali Heshmati [a], Seyed Massoud Houshmand [c,1], Kolsoum InanlooRahatloo [a,1], Forouzandeh Mahjoubi [c,1], Mohammad Hossein Sanati [c,1], Hadi Yari [d,1], Afagh Alavi [e,1], Saeid Amel Jamehdar [f,1], Soroosh Dabiri [g,1], Hamid Galehdari [h,1], Mohammad Reza Haghshenas [i,1], Amir Masoud Hashemian [j,1], Abtin Heidarzadeh [k,1], Issa Jahanzad [l,1], Elham Kheyrani [m,1], Ahmad Piroozmand [n,1], Ali Mojtahedi [o,1], Hadi Razavi Nikoo [p,1], Mohammad Masoud Rahimi Bidgoli [e,1], Nayebali Rezvani [q,1], Mehdi Sepehrnejad [r,1], Arash Shakibzadeh [s,1], Gholamreza Shariati [t,1], Noorossadat Seyyedi [u,1], Seyed MohammadSaleh Zahraei [h,1], Iman Safari [a,**], Elahe Elahi [a,*]

[a] School of Biology, University College of Science, University of Tehran, Tehran, Iran
[b] Cell and Molecular Biology Department, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[c] Department of Medical Biotechnology, National Institute for Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran
[d] Human Genetics Division, Medical Biotechnology Department, National Institute of Genetics Engineering and Biotechnology, Tehran, Iran
[e] Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
[f] Antimicrobial Resistance Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
[g] Department of Laboratory Sciences, Zahedan University of Medical Sciences, Zahedan, Iran
[h] Department of Genetics, Faculty of Sciences, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[i] Molecular and Cell Biology Research Center, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran
[j] Department of Emergency Medicine, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[k] Faculty of Medicine, Guilan University of Medical Sciences, Guilan, Iran
[l] Pars Hospital Laboratory, Rasht, Iran
[m] Taban Genetics Laboratory, Tehran, Iran
[n] Department of Microbiology and Immunology, School of Medicine, Kashan University of Medical Sciences, Kashan, Iran
[o] Microbiology Department, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran
[p] Department of Microbiology, Faculty of Medicine, Golestan University of Medical Sciences, Gorgan, Iran
[q] Department of Clinical Biochemistry, Kermanshah University of Medical Sciences, Kermanshah, Iran
[r] COVID-19 Diagnosis Laboratory, Kashan University of Medical Sciences, Kashan, Iran
[s] Department of Medical Virology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
[t] Department of Medical Genetics, Faculty of Medicine, Jundishapur University of Medical Sciences, Ahvaz, Iran
[u] Diagnostic Laboratory Sciences and Technology Research Center, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

ABSTRACT

We aimed to describe SARS-CoV-2 strains in Iranians from nine distributed cities infected during two months expanding late 2020 and early 2021 by genotyping known informative single nucleotide in five PCR amplicons. Two variants associated with haplotype H1 (clade G) and nine additional variants associated with other haplotypes were genotyped, respectively, in RNA isolates of 244 and 85 individuals. The variants associated with the H1a (GR) and H1b (GH) haplotypes were most prevalent, indicating a significant change in infection pattern with passage of time. The most important findings were that recombinant genomes and co-infection, respectively, were surmised in 44.7% and 12.9% of the samples extensively genotyped. Partners of many of the recombinations were relatively common strains. Co-existing viruses were among those currently circulating in Iran. In addition to random mutations, co-infection with different existing strains and recombination between their genomes may significantly contribute to the emergence of new SARS-CoV-2 strains.

* Corresponding author. University College of Science, University of Tehran, College of Science, University of Tehran, Enghelab Ave., Tehran, 1417614411, Iran.
** Corresponding author. University College of Science, University of Tehran, College of Science, University of Tehran, Enghelab Ave., Tehran, 1417614411, Iran.
E-mail addresses: imansafari@ut.ac.ir, iman.safari15@gmail.com (I. Safari), elaheelahi@ut.ac.ir, elahe.elahi@gmail.com (E. Elahi).
1 These authors contributed equally.

1. Introduction

The first SARS-CoV-2 infected individuals from China were described in December 2019, and the first SARS-CoV-2 genome sequence (Wuhan-Hu-1; https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2) retrieved from an individual infected during that period was published on February 3, 2020 (Wu et al., 2020; Zhou et al., 2020). Rapid person to person transmission of the virus resulted in increased numbers of infections in countries throughout the world, and the designation of the disease caused by the virus as the coronavirus disease 2019 (COVID-19) pandemic by the World Health Organization (WHO) on March 11, 2020 (Cucinotta and Vanelli, 2020). Meanwhile, more SARS-CoV-2 genomes pertaining to infected individuals from various countries/territories, but not all, were sequenced and the sequences were submitted to public databases. It is expected that the dearth of whole genome sequence data from some countries/territories was related to cost and technical limitations. This is the case for Iran, a country in which infected individuals were reported in early February 2020, soon after the start of the pandemic (Yavarian et al., 2020). By November 21, 2020 when the present study was initiated, there were only 18 complete high coverage SARS-CoV-2 genome sequences from Iran available at GISAID (Global Initiative on Sharing All Influenza Data; https://plarform.gisaid.org) (supplementary Table S1).

Early studies grouped available SARS-CoV-2 genome sequences from throughout the world into haplotype groups (Forster et al., 2020; Pachetti et al., 2020; Tang et al., 2020). In an extension on these studies, after analysis of all 2790 SARS-CoV-2 genome sequences from 56 countries/territories available at the GISAID database pertaining to viruses isolated from patients prior to the end of March 2020, we described 66 haplotypes (Safari et al., 2021a). The vast majority of the 2790 sequences were associated with these. Each haplotype was defined by the co-segregation of two or more nucleotide sequence variations. The haplotypes were distributed in 13 major haplotype groups, H1–H13; the majority of the sequences were associated with H1–H3. Several major haplotypes had one or more sub-haplotype, each of which had specific distinguishing variation(s) in addition to the defining sequence variations of the major haplotype group. For example, H1a and H1b were two common sub-haplotypes of H1. Later, analysis of 74,992 sequences from individuals infected between June 1, 2020 and November 15, 2020 led to the discovery of rapid expansion of a new haplotype (H1r) that had evolved during this interval (Safari et al., 2021b). The analysis also showed that by mid-November 2020, H1a, H1b, and H1r were the most common haplotypes associated with SARS-CoV-2 sequences available at GISAID.

Others have labeled haplotypes and sub-haplotypes as clades, clusters, or lineages, and the members of the variously labeled groups often correspond (Liu et al., 2020; Rambaut et al., 2020) (GISAID; Nextstrain (https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-naming)). For example, haplotypes H1, H1a, H1b, and H1r correspond, respectively, to clades G, GR, GH, and GV in GISAID, and to clades 20A, 20B, 20C, and 20E in Nextstrain. Here, the haplotype nomenclature in references Safari et al. (2021a and 2021b) will be used as these haplotypes are defined very completely and stringently; all the variations associated with each haplotype that consistently co-segregated were described.

The 18 SARS-CoV-2 genome sequences from Iran that were referred to above pertained to viruses isolated from individuals infected between March 2, 2020 and June 26, 2020. Fifteen of the 18 sequences were associated with the major haplotype group H5 as they had the variant nucleotide at positions 1397 and 28688 (Table S1) (Safari et al., 2021a). Two of the sequences with collection dates in May and June were associated with the major haplotype group H1. We now report the haplotypes associated with genomes of SARS-CoV-2 viruses isolated from more recently infected Iranians (individuals infected in the interval spanning November 21, 2020–January 19, 2021) based on genotypes of Tag and other informative nucleotide variations (Safari et al., 2021a, 2021b). Additionally and importantly, we note evidence for simultaneous presence of more than one virus strain in some infected individuals and multiple instances of recombination between genomes of different SARS-CoV-2 strains.

2. Methods

This research was performed with the approval of the ethics boards of the National Institute of Genetic Engineering and Biotechnology and the University of Tehran.

2.1. RNA samples

In all, 244 RNA isolates from nose and/or throat swabs of individuals confirmed by PCR to be infected with SARS-CoV-2 in the interval between November 21, 2020 and January 19, 2021 were obtained from laboratories of 9 cities in Iran (Supplementary Fig. S1). The cities are the capitals of provinces well dispersed in Iran, and one of the cities, Tehran, is also the capital of the country. All the participating laboratories were authorized by the Ministry of Health and Medical Education of Iran to perform diagnostic tests for SARS-CoV-2 infection. Only one sample from any single family was obtained. Samples were gathered irrespective of age, sex, or the severity of disease presentations of infected individuals.

Fifty RNA isolates from individuals infected during the interval that spans November 21, 2020 and December 20, 2020 (hereafter called Month 9 samples in accordance with the Iranian solar calendar) were obtained from laboratories of each of three cities, Tehran, Rasht, and Ahvaz. Twelve samples from individuals infected during the same time interval were also obtained from each of four additional cities. Subsequently, 46 RNA isolates from individuals infected during the interval between December 21, 2020 and January 19, 2021 (hereafter called Month 10 samples) were obtained. Fourteen were from Tehran, 17 were from Sari which is the capital of the province of Mazandaran, and 15 were from three other cities.

2.2. Genotyping of SARS-CoV-2 genome sequence variations of interest

Five PCR primer pairs were designed such that the amplification products would include positions of previously identified Tag/signature nucleotide variations for various SARS-CoV-2 genome haplotypes or sub-haplotypes (H1, H1a, H1b, and H1r) that had been shown to be commonly associated with SARS-CoV-2 genome sequences retrieved from individuals throughout the world infected by mid-November 2020 and available at GISAID (Table 1, Fig. 1) (Safari et al., 2021a, 2021b). Some amplicons also included other sequence variants of potential interest, including a Tag variation of haplotype H5 that was apparently prevalent in Iran during the early stage of the COVID-19 pandemic (Table S1). The sequences from the early stage of the pandemic from Iran were aligned as described before (Safari et al., 2021a, 2021b). In addition to formal description of the variants, eleven of the variant positions that will be repeatedly referred to are designated V1–V11 for the sake of brevity. The V1–V11 designations correspond to their order in the virus genome (Tables 2–4 and S2). Each RNA was used as template in one or more cDNA synthesis reaction using a cDNA synthesis master mix (BIOFACT, Korea; Cat no.: BR441-096). Some cDNA synthesis reactions contained the reverse PCR primer of each of two primer pairs (Table 1). The cDNAs were subsequently used in routine PCR reactions that contained PCR primer pairs designed for genotyping of the nucleotide variations of interest. The PCR products were sequenced by the Sanger protocol, and sequences were analyzed using Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA). The genome sequence of the Wuhan-Hu-1 isolate (NC_045512.2) was used as the SARS-CoV-2 reference sequence.

Amplicons 1 and 3 were amplified and sequenced in all these samples. These amplicons include Tag single nucleotide variations (SNVs;

Table 1

Sequences of primers used for synthesis of PCR amplicons that include SARS-CoV-2 genome sequence variations of interest[a].

Amplicon no.	Genomic region within amplicon	Primer sequences (5'- 3')[b]	Target Tag SNV/haplotype	Other sequence variations of interest within the amplicon
1	13961–14601	F: TATACGCCAACTTAGGTGAACG	14408C>T/H1	
		R: TAGATTACCAGAAGCAGCGTG		
2	21770–22446	F: GTCTCTGGGACCAATGGTAC	22227C>T/H1r	21991-21993delTTA/B.1.1.7
		R: GGGTCAAGTGCACAGTCTAC		
3	22855–23562	F: CTGCGTTATAGCTTGGAATTCT	23403A>G/H1	23012G>A/B.1.351; 23063A>T/B.1.1.7 & B.1.351; 23271C>A/B.1.1.7
		R: CCAATGGGTATGTCACACTCA		
4	25318–25940	F: CTGCTGCAAATTTGATGAAGAC	25563G>T/H1b	
		R: TCATGTTCAGAAATAGGACTTGT		
5	28497–29161	F: ACACCAATAGCAGTCCAGATG	28881_28883GGG>AAC/ H1a	28688T>C/H5; 28932C>T/H1r; 28977C>T/B.1.1.7
		R: AGTTCCTTGTCTGATTAGTTCCT		

[a] , The cDNAs pertaining to amplicons 1 and 3 were usually synthesized together in a single cDNA synthesis reaction, as were the cDNAs pertaining to amplicons 4 and 5.

[b] Reverse (R) primers were used in all of the cDNA synthesis reactions.
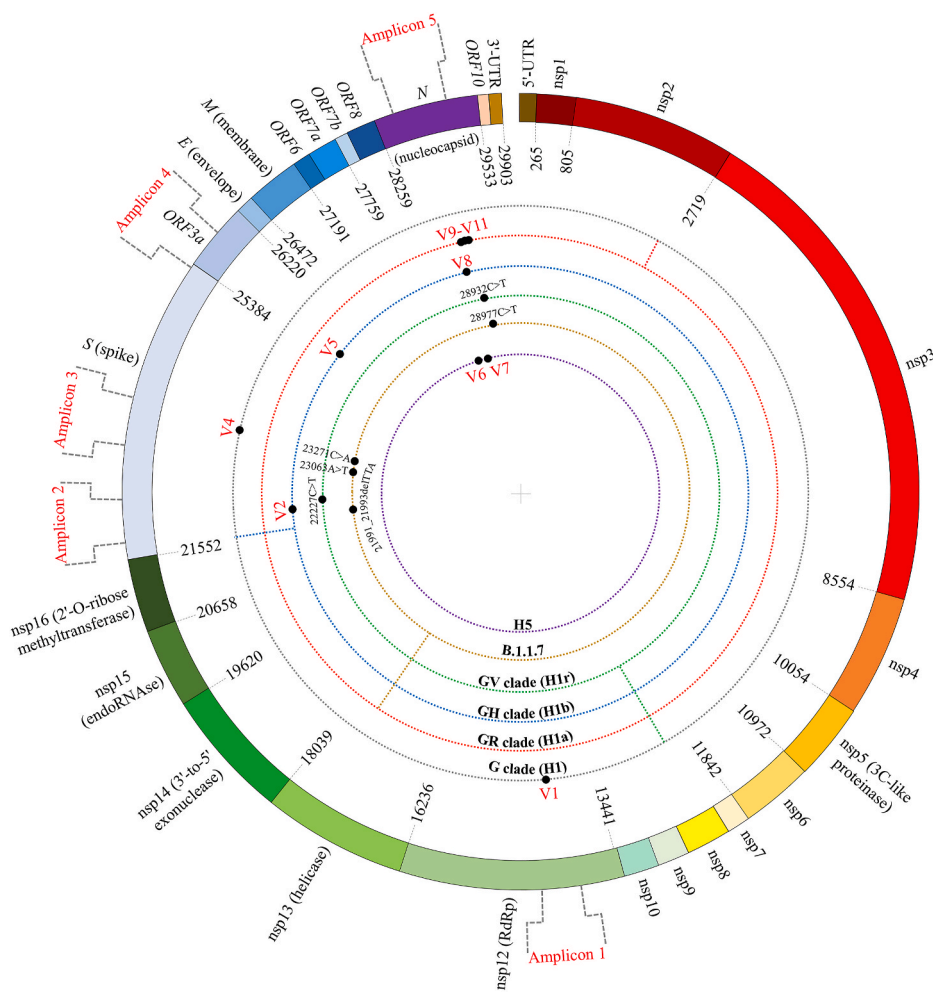

Fig. 1. Schematic presentation of parts of the SARS-CoV-2 genome that were PCR amplified and sequenced. The span of each of the amplicons 1–5 is shown. V1–V11 that are the positions of various informative sequence variations and the positions of six other variations that were genotyped are indicated with filled circles. The virus strains that associate with the sequence variations are also designated. V3 (22992G>A) is not shown because it is observed on both the H1a and H1b backgrounds (see text). The genes of the virus genome are shown, and the names of several of the encoded proteins are written in parentheses.

14408C>T (V1) and 23403A>G (V4)) associated with the major haplotype group H1. Amplicons 4 and 5 were amplified and sequenced in Month 9 samples from Tehran and in four samples from Rasht. Amplicon 5 includes Tag SNVs associated with sub-haplotype H1a (28881_28883GGG>AAC (V9–V10–V11)), sub-haplotype H1r (28932C>T), and the major haplotype group H5 (28688T>C (V6)). A Tag variation associated with sub-haplotype H1b (255563G>T (V5)) is positioned in amplicon 4. A second Tag SNV of H1r (22227C>T) in amplicon 2 was genotyped in a subset of 12 samples from Tehran. Various combinations of amplicons 1–5 in the Month 10 samples were amplified and sequenced. In addition to Tag SNVs associated with

haplotypes H1, H1a, H1b, H1r, and H5, four sequence variations (21991_21993delTTA, 23063A>T, 23271C>A, and 28977C>T) associated with the B.1.1.7 strain that towards the end of our survey was reported to be expanding in the United Kingdom and some other European countries were positioned in the amplicons sequenced. Therefore, these were also genotyped in the Iranian samples.

2.3. Inference of co-infection and recombination

Co-infection with more than one viral strain in an individual was inferred upon visualization of heterozogosity at one or more nucleotide

positions. All samples with heterozygosity were amplified and sequenced at least twice. Some sequences were inferred by visual inspection to be products of recombination events.

3. Results

3.1. Genotyping results of month 9 samples: high frequency of variant nucleotides of H1a (GR) and H1b (GH); evidence of co-infection and recombination

All 198 RNA isolates of Month 9 from seven cities had the V1 and V4 variations associated with haplotype H1 (Tables S2 and 2). Two of the samples (R-1 and R-2) were heterozygous for variation V1, suggesting co-infection with two virus strains (Fig. 2). Among the 50 samples from Tehran, 14 had the variant genotype at V9–V10–V11 which is associated with H1a. One of these (T-3), also had the variant allele 22992G>A (V3). Variant V3 was of interest because it was also observed twice (T-42 and T-18) on an H1b background, whereas Nexstrain includes this variation as a marker for clade 20F that is derived from clade 20B (corresponding to H1a). Thirty-four of the non-H1a Month 9 samples from Tehran had the variant nucleotide at the H1b Tag V5. Therefore, 48 (96% of) the

samples from Tehran had variant nucleotides at H1a or H1b Tag positions, and the ratio of these was 1H1a: 2.4H1b. Neither the H1a nor the H1b variants were seen in two samples (T-47 and T-51). It was observed that a variant nucleotide at position 28854 (V8) in amplicon 5 was strongly linked to the H1b Tag SNV (V5) in the samples from Iran, as only two of the 34H1b samples lacked the variant allele at this position. One of these (T-35) was a "pure" H1b in the sense that only Tag variants of the H1 (V1 & V4) and H1b (V5) haplotypes were observed in the genome. The other (T-4) proved to be more interesting (see below). The Tag variant nucleotide of haplotype H5 (V6) was observed in one sample from Tehran (T-47) and also in four samples from Rasht (R-1, R-2, R-3, and R-4), and this observation also proved to be interesting (see below). The variant nucleotide at position 28835 (V7) was present in all five samples with the V6 variant, indicating possible linkage between the two. R-3 among the samples from Rasht showed heterozygosity at V5, suggesting infection with two viruses in a third sample from this city (Fig. 2). The H1r associated Tag variant at position 28932 was not present in the 54 samples in which amplicon 5 was sequenced. Also, amplification and sequencing of amplicon 2 in 12 samples from Tehran evidenced that none of these had the H1r Tag variant at position 22227. However, a variant three nucleotide deletion 22189_22191delTAT (V2)
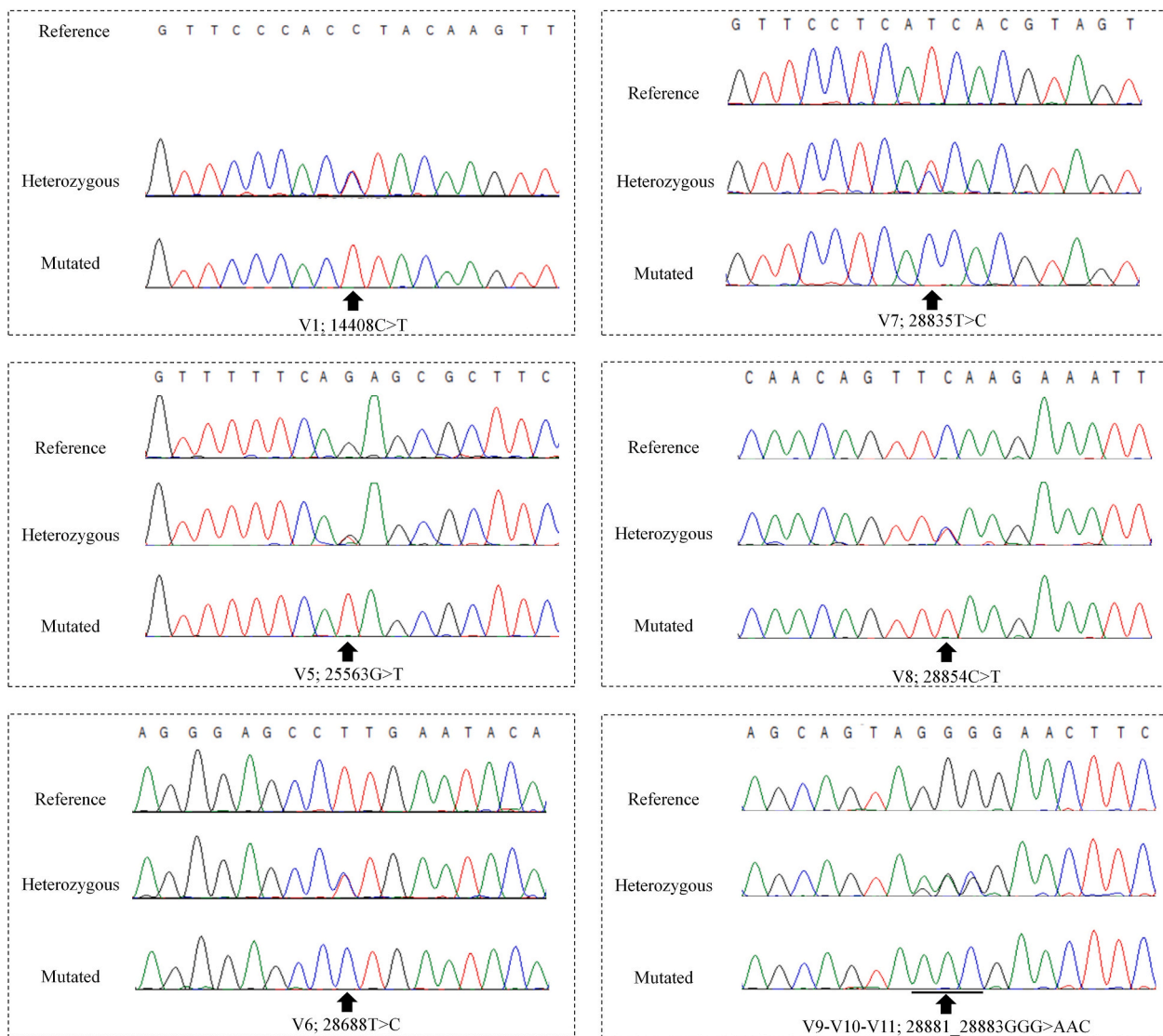
Fig. 2. DNA sequence chromatograms of PCR amplicons of the SARS-CoV-2 genome indicative of co-infection with two virus strains. Chromatograms that include variant positions with sequences that correspond to the reference sequence and a mutated sequence, and chromatograms of apparently heterozygous status that are suggestive of co-infection are shown. A sequence at V1 with the reference sequence was not observed among the samples sequenced.

Table 2

Genotyping of genome sequence variations associated with SARS-CoV-2 haplotypes in viruses isolated from Iranians infected between November 21 and December 20, 2020 (Month 9)[x].

Sequence variation		14408C>T (RdRp: p.Pro 312Leu)	22189_2219 1delTAT (S: p.Ile210del)	22992G>A (S: p.Ser477Asn)	23403A>G (S: p.Asp614Gly)	25563G>T (ORF3a: p.Gln57His)	28688T>C (N: p.Leu 139Phe)	28835T> C (N: p.Ser188Pro)	28854C>T (N: p. Ser194Leu)	28881G>A (N: p.Arg203Lys)	28882G>A (N: p. Arg203Arg)	28883G>C (N: p.Gly 204Arg)
Associated haplotype*/Virus strain*		**H1/G**			**H1/G**	**H1b/GH**	**H5**	**H5 linked**	**H1b linked**	**H1a/GR**	**H1a/GR**	**H1a/GR**
Variation ID		**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10**	**V11**
Genome type#	**No. Samples in Tehran (Total: 50)**											
A (=H1a)	13	+	.		+	.	.	.	.	+	+	+
H1b	1 (T-35)	+	.		+	+	.	.	.	.	.	.
B (=H1b+V8)@	8	+	.	.	+	+	.	.	+	.	.	.
C (=H1b+ V2+V8)@	21	+	+	.	+	+	.	.	+	.	.	.
Tehran samples proposed to be infected with viruses whose genomes are products of recombination												
	No. Samples (ID)											
D	1 (T-47)	+	.	.	+	.	+	+	.	.	.	.
E	1 (T-51)	+	.	.	+	.	.	.	+	.	.	.
F	1 (T-3)	+	.	+	+	.	.	.	.	+	+	+
G	1 (T-42)	+	.	+	+	+	.	.	+	.	.	.
H	1 (T-18)	+	+	+	+	+	.	.	+	.	.	.
I	1 (T-4)	+	+	.	+	.	.	.	.	.	.	.
Rasht samples proposed to be infected with two viruses and/or with viruses whose genomes are products of recombination												
D	1 (R-4)	+	.	.	+	.	+	+	.	.	.	.
D and J	1 (R-3)	+	.	.	+	G/T	+	+	.	.	.	.
D and K	2 (R-1 & R-2)	C/T	.	.	+	.	+	+	.	.	.	.
Other samples from provinces outside of Tehran												
Rasht	46 samples	+			+							
Ahvaz	50 samples	+			+							
Kashan	12 samples	+			+							
Tabriz	12 samples	+			+							
Kermanshah	12 samples	+			+							
Shiraz	12 samples	+			+							

[x], Summary of data presented in Table S2; *, Haplotypes described in Safari et al, (2021a); **, virus strains described in GISAID; #, each type refers to a genome with a unique combination of sequence variations; @, number of sequences with and without the variant nucleotide at position V2 extrapolated on basis of results on 12 samples in which amplicon 2 was amplified and sequenced.

+: variant sequence:., reference sequence, empty cell: not sequenced.

in amplicon 2 was observed in nine of the 12 subset of H1b samples, including the H1b sample T-4 described above that did not have variant V8. By extrapolation, it is expected that the deletion at V2 would be present in approximately three fourth of the H1b sequences from Tehran. (The deletion at V2 causes an in frame deletion p. Ile210del in the spike protein.) Variant alleles at the two SNVs within amplicon 3 that are associated with strain B.1.1.7 were not observed in the 198 Month 9 samples in which this amplicon was sequenced, and the linked nucleotide at the third B.1.1.7 marker within amplicon 5 was not seen in 54 samples from Tehran and Rasht. The data suggest that the H1r and B.1.1.7 stains did not have a significant presence in Iran during Month 9.

Among the 54 samples (50 from Tehran and four from Rasht) that were genotyped more extensively than the other Month 9 samples, A, B, and C were the most frequent genome types. A corresponds to H1a. B and C presumably evolved by addition of one or two mutations on H1b backgrounds. Genome type C may correspond to the very recently described B.1.36 lineage of the Phylogenetic Assignment Named Global Outbreak (PANGO) Lineages (GISAID). Like genome type C, B.1.36 is defined by the variant nucleotides at V1, V2, V4, V5, and V8. The sequence data pertaining to ten samples (18.5%) suggested infection with viruses that had recombinant genomes (Table 2). Parsimony

models for the recombination events that produced the presumed recombinant genomes are presented in Fig. 3. For each putative recombinant genome, at least one possible driving recombination event is described in which one or both partners have genome types that are among common Month 9 genome types (genome types A and C) or was apparently common among genome sequences from Iran during earlier stages of the pandemic (H5) (Table 2 and S1, Fig. 3). Even the uncommon genome types that are proposed recombination partners were each independently recorded at least once in the sequences reported here. Recombination was surmised for the proposed recombinant genomes because of co-existence of sequence variations that were previously established to be Tag or tightly linked nucleotides of various haplotypes. Consistent with some mosaic genomes being products of recombination, most of the possibly recombinant genomes had long stretches of thousands of nucleotides defined by multiple variant nucleotides associated with the respective partners of the respective recombination events. The possibility of recurrent mutation rather than recombination as the origin of these sequences seems unlikely precisely because the co-existing variations were not random.
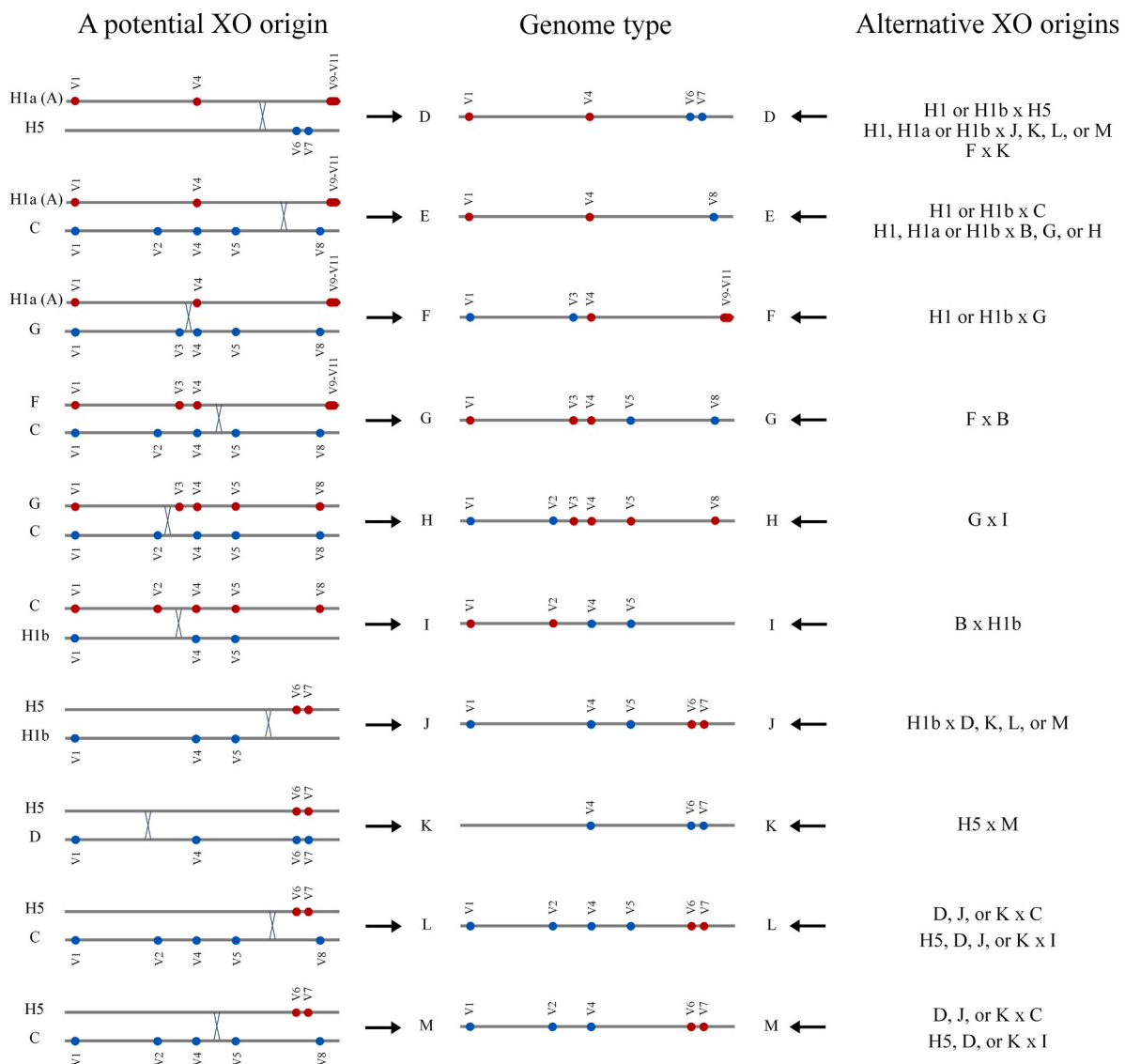

Fig. 3. Schematic presentation of potential recombination events between genomes of SARS-CoV-2 strains that produced putative observed recombinant genomes. XO, cross over.

Table 3

Genotyping of genome sequence variations associated with SARS-CoV-2 haplotypes in viruses isolated from Iranians (Tehran, Mashhad, Ahvaz, Shiraz) infected between December 21, 2020 and January 19, 2021 (Month 10)[x].

Sequence variation		14408C>T (RdRp: p. Pro312Leu)	22189_22191delTAT (S: p.Ile210del)	22992G>A (S: p. Ser477Asn)	23403A>G (S: p. Asp614Gly)	25563G>T (ORF3a: p. Gln57His)	28688T>C (N: p. Leu139Phe)	28835T>C (N: p. Ser188Pro)	28854C>T (N: p. Ser194Leu)	28881G>A (N: p. Arg203Lys)	28882G>A (N: p. Arg203Arg)	28883G>C (N: p. Gly204Arg)
Associated haplotype*/ Virus strain		**H1/G**			**H1/G**	**H1b/GH**	**H5**	**H5 linked**	**H1b linked**	**H1a/GR**	**H1a/GR**	**H1a/GR**
Variation ID		**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10**	**V11**
Genome type[#]	**No. Samples in Tehran (Total: 14)**											
D	4	+	.	.	+	.	+	+	.	.	.	.
L	3	+	+	.	+	+	+	+	.	.	.	.
M	2	+	+	.	+	.	+	+	.	.	.	.
Tehran samples proposed to be infected with two viruses	**No. Samples (ID)**											
D and H1	1 (T-30)	+	.	.	+	.	T/C	T/C	.	.	.	.
A and D	1 (T-20)	+	.	.	+	.	T/C	T/C	.	G/A	G/A	G/C
C and L	1 (T-13)	+	+	.	+	+	T/C	T/C	C/T	.	.	.
M and L	1 (T-18)	+	+	.	+	G/T	+	+	.	.	.	.
M and N	1 (T-26)	+	+	.	+	.	T/C	T/C	.	.	.	.
Samples from provinces outside of Tehran												
Mashhad	4 samples	+	+	.	+							
Mashhad	1 samples	+	.	.	+							
Ahvaz	5 samples	+	+	.	+							
Shiraz	5 samples	+	.	.	+							

[x], Summary of data presented in Table S2; *, Haplotypes described in Safari et al, (2021a); **, virus strains described in GISAID; #, each type refers to a genome with a unique combination of sequence variations. +: variant sequence:., reference sequence, empty cell: not sequenced.

Table 4

Genotyping of SARS-CoV-2 genome sequence variations of 17 virus samples from Sari collected between December 21, 2020 and January 19, 2021 (Month 10)[x].

Sequence variation		14408C>T (RdRp: p.Pro312Leu)	22189_22191delTAT (S: p.Ile210del)	22992G>A (S: p.Ser477Asn)	23403A>G (S: p.Asp614Gly)	25563G>T (ORF3a: p.Gln57His)	28688T>C (N: p.Leu139Phe)	28835T>C (N: p.Ser188Pro)	28854C>T (N: p.Ser194Leu)	28881G>A (N: p.Arg203Lys)	28882G>A (N: p.Arg203Arg)	28883G>C (N: p.Gly204Arg)
Associated haplotype*/Virus strain**		H1/G		H1/G	H1/G	H1b/GH	H5	H5 linked	H1b linked	H1a/GR	H1a/GR	H1a/GR
Variation ID		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
Genome type#	No. Of samples											
H1b	1	+		.	+	+	.	.	.	.	.	.
B or C	2	+		.	+	+	.	.	+	.	.	.
D	5	+		.	+	.	+	+	.	.	.	.
F	1	+		+	+	.	.	.	.	+	+	+
J	5	+		.	+	+	+	+	.	.	.	.
Samples proposed to be infected with two viruses												
A and D	3	+		.	+	.	T/C	T/C	.	G/A	G/A	G/C

X, Summary of data presented in Table S2; *, Haplotypes described in Safari et al., (2021a); **, virus strains described in GISAID; #, each type refers to a genome with a unique combination of sequence variations. +: variant sequence., reference sequence, empty cell: not sequenced.

3.2. Genotyping results of month 10 samples: increased frequencies of co-infection and recombination

As was the case for the Month 9 samples, all 46 Month 10 samples had the V1 and V4 variant nucleotides associated with the H1 major haplotype (Tables 3 and 4). Also, nucleotide variations associated with the H1r and B.1.1.7 strains were not present in these samples. Other than these commonalities, the genotype patterns of Month 10 and Month 9 samples were different. It seems unlikely that sample size alone could account for these differences. Whereas over 25% (14/50) of Month 9 samples from Tehran had the variant genotype at the H1a Tag V9–V10–V11, only one among 14 Month 10 samples from Tehran had this variant genotype (present in T-20 simultaneously with virus of genome type D) (Table 3). (The variant genotype at the H1a Tag V9–V10–V11was also absent in seven additional Month 10 samples from Tehran in which this Tag was specifically genotyped (not shown)). The H1b Tag variant (V5) that had been observed in 68% (34/50) of Month 9 Tehran samples, was found in five (36%) Month 10 Tehran samples. Four (29%) Month 10 Tehran samples did not have the variant sequence at either H1a (V9–V10–V11) or H1b Tag (V5) nucleotide positions. The comparable figure for Month 9 Tehran sequences was 4% (2/50). Presumably recombinant genome types (D, L, M, and/or N) were present in all 14 Month 10 Tehran samples, and five of the 14 samples were simultaneously infected with more than one virus strain (Table 3, Figs. 2 and 3). The common Month 9 genome types A and C and haplotype H5 are potential partners in the recombination events that produced genome types D, L, and M (Table 3, Table S1, Fig. 3).

In addition to Tehran, the Month 10 samples from Sari were extensively genotyped (Table 4). H1a or H1b Tag variations were present in 70.6% (12/17) of the sequences from this city. Two of the Month 10 samples from Sari contained the B or C genome type that had been observed among the Month 9 samples of Tehran. Three samples contained the A genotype, and all three of these samples were simultaneously infected with viruses that had the recombinant type D genome. In fact, most (14/17 = 82.4%) Month 10 samples of from Sari contained recombinant genomes. All these putative recombinant genome types (D, F, and J) had previously been described among the Month 9 samples from Tehran.

3.3. Summary of findings on recombinant genomes

With respect to recombinant genomes and co-infection with at least two virus strains, the findings pertaining to the 85 samples of Month 9 and Month 10 that were more extensively genotyped can be summarized as follows. Thirty-eight (44.7%) samples were infected with viruses with possibly recombinant genomes. The three genome types A, C, and H5 were relatively common partners of the proposed recombination events. As these were in Month 9 or in the past common genome types in Iran, it is expected that they would be more likely partners of recombination events. Genome type A with haplotype H1a (clade GR) is a world-wide common strain. Genome type C is closely related to the H1b haplotype (clade GH) that is also a world-wide common strain. H5 that was apparently commonly associated with sequences from Iran in the early stages of the COVID-19 pandemic, was also the most frequent haplotype among the low number of sequences that were available from other countries of the Middle East during the same period (Safari et al., 2021a). Other partners (D, G, F, and I) of putative recombination events may themselves be products of local recombination incidences.

4. Discussion

It is well known that the 23403A>G variation in the *S* gene that is a Tag variation for haplotype H1 is present in the vast majority of available SARS-CoV-2 genome sequences that were isolated from individuals world-wide who were infected with the virus from July 2020 onwards (https://cov.lanl.gov) (Safari et al., 2021b). We have now critically

shown that this variation and the associated major haplotype H1 are also highly prevalent in the genomes of virus isolates from individuals of various regions of Iran who were infected between November 2020 and January 2021. This signifies a shift from earlier prevalence of the H5 haplotype. Analysis at the sub-haplotype level proved to be more complicated because of putative recombination events that created novel genome types and because of instances of co-infection with two virus strains. Even though sample sizes were not very large, the data that are described in detail in the Results section do suggest that the allele frequencies of various informative genome positions associated with sub-haplotypes may change within the short time span of one month. For example, the frequencies of the variant nucleotides at the H1a Tag V9–V10–V11 among Month 9 and Month 10 samples from Tehran were significantly different. Whereas universal overtake of the 23403A>G variation has been attributed to biological parameters associated with this variation including virus load and possibly infectivity, we suspect that the changes observed here are most likely due to demographic behavior patterns and stochastic events (Gobeil et al., 2021; Korber et al., 2020b; Plante et al., 2020; Volz et al., 2021; Yurkovetskiy et al., 2020).

The most important findings of this study are the frequent observation of genomes that are possibly the products of recombination events between the genomes of different SARS-CoV-2 strains, and the simultaneous presence of more than one virus strain in some infected individuals (see postscript). Recombination occurs in many RNA viruses, more frequently in those with positive-sense genomes (Simon-Loriere and Holmes, 2011). The contribution of recombination to the evolution of Betacoronaviruses is well established (Bobay et al., 2020; Graham and Baric, 2010; Lai and Cavanagh, 1997; Su et al., 2016). Among the coronaviruses, the recombinatory origins of the clinically important human SARS-CoV, MERS-CoV, and 2019 SARS-CoV-2 are well documented (Hon et al., 2008; Li et al., 2020; Wang et al., 2015; Zhang et al., 2016; Zhu et al., 2020). And, despite its relatively recent emergence, there already exists evidence that recombination between co-existing SARS-CoV-2 strains may contribute to the genome diversity of this virus (Giorgi et al., 2021). In fact, recombination events may have contributed to the origin of some common haplotypes or strains (e.g. strain B.1.1.7) that are defined by co-segregation of multiple sequence variations. Of course, it is possible that such complicated haplotypes may have evolved by accumulations of sequential mutations and subsequent expansion due to stochastic events or selection. A more probable scenario is that these haplotypes are products of one or more recombination event between virus strains that each contained a subset of the variations. To the best of our knowledge, the first report of recombination between SARS-CoV-2 strains was published in August of 2020 (Yi, 2020). Recombination in that study was deduced on the basis of haplotype network analysis. In a few subsequent studies, recombination events between SARS-CoV-2 strains have been surmised on the basis defining markers of major clades or sequence variations in locally circulating strains (Korber et al., 2020a; VanInsberghe et al., 2020; Varabyou et al., 2020). The distinguishing feature of our results is that a relatively high fraction (38/85; 44.7%) of the samples surveyed contained recombinant genomes. Technical issues in sequence analysis in other studies may have resulted in underestimation of recombinant genomes (Varabyou et al., 2020). The detailed scrutinizing visual approach applied in this study readily allowed recognition of potential recombination events; this approach is not suitable for analysis of large numbers of sequences and has presumably not been applied. Finally, it is evident that detection of recombination depends on co-existence of different strains within the same infected individual. Co-existence of different strains is more likely with the evolvement of more novel strains and as infection rates increase. With the passage of time, novel strains have increasingly evolved and infection rates have increased. For these reasons, recombinant genomes may have been less common in the past.

Till now, there have only been a few reports of co-infection with two SARS-CoV-2 strains (Baang et al., 2021; Choi et al., 2020; da Silva Francisco Jr et al., 2021; Pedro et al., 2021). The co-infections in these cases were proposed to demonstrate within host evolution of the virus or independent infections. It is notable that our sequencing data suggest the presence of two virus strains in 11 of the 85 (12.9%) samples that were relatively more extensively genotyped. The sequencing data is consistent with the proposal that the co-existing viruses are among those circulating in Iran during Months 9 and 10. A virus with the H5 haplotype was not present in the co-infected samples, suggesting that the frequency of this haplotype has significantly decreased. The relatively frequent A and C genome types among Month 9 samples of Tehran were present in five of the eleven co-infected samples. The presumably locally evolved genome types D, G, J, K, L, and/or M genome types were identified in all of the co-infected samples. Of course, the putative recombinant genome types D-J, L, and M that were observed in some samples infected with a single virus must themselves have evolved in cells that were co-infected with two virus strains. Again, it is possible that technical issues in sequence analysis may have resulted in not detecting some instances of co-infection in previous studies. For example, heterozygosity at various positions that would reflect co-infection may be reported as ambiguous reads (N). Also, as was suggested for recombination, increases in present rates of infection may have increased the likelihood of co-infection and thus facilitated identification of co-infected samples.

In light of the fact that new strains of SARS-CoV-2 may have changed pathogenicity and because new strains must be taken into consideration in the framework of design and testing of efficacy of vaccines, drugs, and diagnostic tools, it is important to recognize that in addition to random mutations, co-infection with different existing strains and recombination between their genomes may contribute to the emergence of new strains (Faria et al., 2021; Korber et al., 2020b; Tang et al., 2021; Tegally et al., 2020). Our study evidences that in lieu of whole genome sequencing, genotyping informative sequence variations can be useful for identification of new strains and for describing the distribution of infecting virus strains in a population. This approach was here used for Iran.

Postscript: Due to technical and financial issues, whole genome sequencing was not performed on the samples screened in this study (see Introduction). Fortunately, limitations have been partially alleviated in the interim after the reported study, and 110 SARS-CoV-2 complete whole genome sequences from Iran were available at GISAID on June 1, 2021. These sequences may serve as an acceptable resource for assessment of presence and frequency of possible recombinant genomes based on data of whole genome sequences (in lieu of variant nucleotide genotyping by PCR and Sanger sequencing as in our study). Of the 25 samples collected prior to May 1, 2020, 23 had a pure H5 genome type, consistent with the prevalence of that genotype in Iran during the early phase of the pandemic. Because of lack of heterogeneity, recombination events in this phase of the pandemic would not have been recognized. Of the remaining 85 sequences, 24 and 10, respectively, were associated with the B.1.1.7 and B.1.617+ lineages. All these samples were collected and genotyped in 2021 precisely because the infected individuals had recently been in the UK or India. Therefore, recombination was sought among the 51 remaining sequences. Based on co-existence of sequence variations that were previously established to be Tag or tightly linked nucleotides of various haplotypes or clades, thirteen potentially recombinant genomes were observed among these 51 sequences (25.5%). This estimate is consistent with the relatively high frequency of possible recombinant genomes that was estimated among the Month 9 and Month 10 samples. The genome type of nine of the thirteen recombinant genomes (GISAID IDs: EPI_ISL_514753, EPI_ISL_959279, EPI_ISL_862075, EPI_ISL_862081, EPI_ISL_1014686, EPI_ISL_2254719, EPI_ISL_2253220, EPI_ISL_2227281, and EPI_ISL_2227280) were reported among the Month 9 and Month 10 samples, and the genome types of four (EPI_ISL_1398364, EPI_ISL_815256, EPI_ISL_1014678, and EPI_ISL_2254498) were not earlier observed.

Data availability statement

The data that support the finding of this study are available from the corresponding authors upon reasonable request.

Funding

This research was supported by a grant from the National Institute for Genetic Engineering and Biotechnology, Tehran, Iran.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Peyman Taghizadeh: Investigation, Methodology, Validation. **Sadegh Salehi:** Investigation, Methodology, Validation. **Ali Heshmati:** Investigation, Methodology, Validation. **Seyed Massoud Houshmand:** Resources, Funding acquisition. **Kolsoum InanlooRahatloo:** Methodology. **Forouzandeh Mahjoubi:** Resources, Funding acquisition. **Mohammad Hossein Sanati:** Resources, Funding acquisition. **Hadi Yari:** Resources. **Afagh Alavi:** Resources. **Saeid Amel Jamehdar:** Resources. **Soroosh Dabiri:** Resources. **Hamid Galehdari:** Resources. **Mohammad Reza Haghshenas:** Resources. **Amir Masoud Hashemian:** Resources. **Abtin Heidarzadeh:** Resources. **Issa Jahanzad:** Resources. **Elham Kheyrani:** Resources. **Ahmad Piroozmand:** Resources. **Ali Mojtahedi:** Resources. **Hadi Razavi Nikoo:** Resources. **Mohammad Masoud Rahimi Bidgoli:** Resources. **Nayebali Rezvani:** Resources. **Mehdi Sepehrnejad:** Resources. **Arash Shakibzadeh:** Resources. **Gholamreza Shariati:** Resources. **Noorossadat Seyyedi:** Resources. **Seyed MohammadSaleh Zahraei:** Resources. **Iman Safari:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Elahe Elahi:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

We acknowledge and thank Doctors Ehsan Arefian, Hassan Behboudi, Omid Mirshamsi, Mehrnaz Narooie-Nejad, and Gholamreza Zarrini for facilitating sample collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virol.2021.06.004.

References

Baang, J.H., Smith, C., Mirabelli, C., Valesano, A.L., Manthei, D.M., Bachman, M.A., Wobus, C.E., Adams, M., Washer, L., Martin, E.T., 2021. Prolonged severe acute respiratory syndrome coronavirus 2 replication in an immunocompromised patient. J. Infect. Dis. 223, 23–27.

Bobay, L.-M., O'Donnell, A.C., Ochman, H., 2020. Recombination events are concentrated in the spike protein region of Betacoronaviruses. PLoS Genet. 16, e1009272.

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., Solomon, I.H., Kuo, H.-H., Boucau, J., Bowman, K., 2020. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. N. Engl. J. Med. 383, 2291–2293.

Cucinotta, D., Vanelli, M., 2020. WHO declares COVID-19 a pandemic. Acta bio-medica Atenei Parm. 91, 157–160.

da Silva Francisco Jr., R., Benites, L.F., Lamarca, A.P., de Almeida, L.G.P., Hansen, A.W., Gularte, J.S., Demoliner, M., Gerber, A.L., de C Guimarães, A.P., Antunes, A.K.E., 2021. Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. Virus Res. 296, 198345.

Faria, N.R., Claro, I.M., Candido, D., Moyses Franco, L.A., Andrade, P.S., Coletti, T.M., Silva, C.A.M., Sales, F.C., Manuli, E.R., Aguiar, R.S., 2021. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological.Org 1–9.

Forster, P., Forster, L., Renfrew, C., Forster, M., 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc. Natl. Acad. Sci. Unit. States Am. 117, 9241–9243.

Giorgi, E.E., Bhattacharya, T., Fischer, W.M., Yoon, H., Abfalterer, W., Korber, B., 2021. Recombination and Low-Diversity Confound Homoplasy-Based Methods to Detect the Effect of SARS-CoV-2 Mutations on Viral Transmissibility. bioRxiv.

Gobeil, S.M.-C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., Stalls, V., Kopp, M.F., Henderson, R., Edwards, R.J., 2021. D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. Cell Rep. 34, 108630.

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84, 3134–3146.

Hon, C.-C., Lam, T.-Y., Shi, Z.-L., Drummond, A.J., Yip, C.-W., Zeng, F., Lam, P.-Y., Leung, F.C.-C., 2008. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. J. Virol. 82, 1819–1826.

Korber, B., Fischer, W., Gnanakaran, S.G., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E.E., Bhattacharya, T., Parker, M.D., 2020a. 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.

Lai, M.M.C., Cavanagh, D., 1997. The molecular biology of coronaviruses. Adv. Virus Res. 48, 1–100.

Li, X., Giorgi, E.E., Marichannegowda, M.H., Foley, B., Xiao, C., Kong, X.-P., Chen, Y., Gnanakaran, S., Korber, B., Gao, F., 2020. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. 6, eabb9153.

Liu, S., Shen, J., Fang, S., Li, K., Liu, J., Yang, L., Hu, C.-D., Wan, J., 2020. Genetic spectrum and distinct evolution patterns of SARS-CoV-2. Front. Microbiol. 11, 2390.

Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R.C., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J. Transl. Med. 18, 1–9.

Pedro, N., Silva, C.N., Magalhães, A.C., Cavadas, B., Rocha, A.M., Moreira, A.C., Gomes, M.S., Silva, D., Sobrinho-Simões, J., Ramos, A., 2021. Dynamics of a dual SARS-CoV-2 lineage Co-infection on a prolonged viral shedding COVID-19 case: insights into clinical severity and disease duration. Microorganisms 9, 300.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., 2020. Spike mutation D614G alters SARS-CoV-2 fitness. Nature 1–6.

Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407.

Safari, I., InanlooRahatloo, K., Elahi, E., 2021a. Evolution of SARS-CoV-2 genome from December 2019 to late March 2020: emerged haplotypes and informative Tag nucleotide variations. J. Med. Virol. n/a. 93 (4), 2010–2020. https://doi.org/10.1002/jmv.26553.

Safari, I., InanlooRahatloo, K., Elahi, E., 2021b. World-wide tracking of major SARS-CoV-2 genome haplotypes in sequences of June 1 to November 15, 2020 and discovery of rapid expansion of a new haplotype. J. Med. Virol. 93 (5), 3251–3256. https://doi.org/10.1002/jmv.26802.

Simon-Loriere, E., Holmes, E.C., 2011. Why do RNA viruses recombine? Nat. Rev. Microbiol. 9, 617–626.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C.K., Zhou, J., Liu, W., Bi, Y., Gao, G.F., 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol. 24, 490–502.

Tang, J.W., Tambyah, P.A., Hui, D.S.C., 2021. Emergence of a new SARS-CoV-2 variant in the UK. J. Infect. 82 (4), e27–e28. https://doi.org/10.1016/j.jinf.2020.12.024.

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., 2020. On the origin and continuing evolution of SARS-CoV-2. Natl. Sci. Rev. 7 (6), 1012–1023. https://doi.org/10.1093/nsr/nwaa036.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., 2020. Emergence and Rapid Spread of a New Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Lineage with Multiple Spike Mutations in South Africa. medRxiv.

VanInsberghe, D., Neish, A.S., Lowen, A.C., Koelle, K., 2020. Identification of SARS-CoV-2 Recombinant Genomes. BioRxiv.

Varabyou, A., Pockrandt, C., Salzberg, S.L., Pertea, M., 2020. Rapid Detection of Inter-clade Recombination in SARS-CoV-2 with Bolotie. Biorxiv.

Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., 2021. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. Cell 184, 64–75.

Wang, Y., Liu, D., Shi, W., Lu, R., Wang, W., Zhao, Y., Deng, Y., Zhou, W., Ren, H., Wu, J., 2015. Origin and possible genetic recombination of the Middle East respiratory

syndrome coronavirus from the first imported case in China: phylogenetics and coalescence analysis. mBio 6.

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.

Yavarian, J., Shafiei-Jandaghi, N.-Z., Sadeghi, K., Malekshahi, S.S., Salimi, V., Nejati, A., Aja-Minejad, F., Ghavvami, N., Saadatmand, F., Mahfouzi, S., 2020. First cases of SARS-CoV-2 in Iran, 2020: case series report. Iran. J. Public Health 49, 1564.

Yi, H., 2020. 2019 novel coronavirus is undergoing active recombination. Clin. Infect. Dis. 71, 884–887.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., 2020. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell 183, 739–751.

Zhang, Z., Shen, L., Gu, X., 2016. Evolutionary dynamics of MERS-CoV: potential recombination, positive selection and transmission. Sci. Rep. 6, 1–10.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273.

Zhu, Z., Meng, K., Meng, G., 2020. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. Sci. Rep. 10, 1–10.