Identification of novel mutations in SARS-COV-2 isolates from Turkey

Shazia Rehman (shaziarehman7@gmail.com)
Rawalpindi Women University
Tariq Mahmood
Quaid-i-azam University, Islamabad
Ejaz Aziz
GDC Khanpur, Haripur, Pakistan
Riffat Batool
PMAS-UAAR, Rawalpindi, Pakistan

Research Article

Keywords: SARS-CoV-2, S protein, Nsp2, Nsp3, NSP4, NSP12/RdRP, Structural alterations Vaccine engineering

Posted Date: August 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-51012/v1

Version of Record: A version of this preprint was published at Archives of Virology on October 6th, 2020.
See the published version at https://doi.org/10.1007/s00705-020-04830-0.

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), originally emerged from Wuhan, has caused an unprecedented worldwide pandemic in the first half of 2020. Since the first report of SARS-CoV-2 on March 10th, 2020 in Turkey, more than 150,000 people in the country have been infected with this virus. In this study, a total of 80 genomic virulent strains from Turkey which were uploaded in NCBI and GISAID database were analyzed with other genomic sequences from different countries with the aim to characterize notable genomic features of SARS-CoV-2 and to identify some novel mutations. Consistent with other studies, the combination of variants at positions C3037T, C14408T and A23403G were most common mutations (73%), that exist together in isolates from Turkey. Our secondary structure prediction analysis also highlighted 11 unique non-substitutional mutations from viral SARS-COV-2 isolates of Turkey in different regions such as in spike (S) protein and non-structural proteins (Nsp2, Nsp3, NSP4, and NSP12/RdRP). Of these 11 mutations, nine of them have been found to be involved in structural alterations at different sites. 3/9 mutants (A771V, T1238I and G1251V) cause alteration in structure of S protein, while the rest of them induces structural changes in Nsp2 (A206T, R207C, T265I), Nsp3 (A1824V), Nsp4 (M2796I) and NSP12 (A4489V). These mutations identified here might have significant functional implications that needs to be addressed for future studies in the context of vaccine engineering and therapeutic interventions. Moreover, transmission and phylogenetic analysis revealed multiple independent sources of introductions for infection of hCovs in Turkey and close phylogenetic relationship of Turkish strains with Saudi strains.

Introduction

The first case of novel coronavirus outbreak in human was reported in Wuhan (China), was named as COVID-19 by WHO and SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) by International Committee on Taxonomy of Viruses [1]. Since the emergence of COVID-19, based on WHO report on 8th July 2020 more than 11,591,595 confirmed cases were reported from 147 countries along with 537,859 death cases due to rapidly spreading SARS- CoV-2 [2]. Coronavirus is usually spread in birds, humans and other mammals and can lead to severe diseases including intestinal, liver, nervous and respiratory systems. The prevalence of human coronaviruses (hCoVs) was first identified in 1960s. In different studies, seven common human coronaviruses were reported, such as MERS-CoV (Middle East respiratory syndrome coronavirus), CoV-HKU1 (beta coronavirus), CoV-OC43 (beta coronavirus), severe acute respiratory syndrome coronavirus (SARS-CoV), CoV-NL63 (alphacoronavirus), CoV-229E (alphacoronavirus) and recent SARS-CoV-2. CoV-OC43 and CoV-229E are responsible for common cold in adults while CoV-NL63 and CoV-HKU1 are linked with chronic pneumonia and common cold. The other three human coronaviruses (SARS-CoV, MERS-CoV and SARS-CoV-2) caused severe respiratory disorders and are responsible for deaths of thousands of people around the globe [3].

Coronavirus belongs to family Coronaviridae and the order Nidovirales is a single positive strand RNA enveloped virus. SARS-CoV-2 belongs to β-coronavirus. The genome sequence homology suggests that newly emerged SARS-CoV-2 is highly similar with bat SARS-CoV [4]. SARS-CoV-2 and SARS-CoV uses ACE2 (angiotension-converting enzyme 2) as receptor. It has been reported that SARS-CoV-2 generally identifies the ACE2 receptor on the target cell by the S protein for incidence of infection [5]. The general symptoms of COVID-19 are cough, fever and fatigue. However, higher level of creatinine kinase, lactate dehydrogenase and c-reactive protein is also associated with COVID-19 infection [6]. Guan et al. [6] reported that the transmission power of SARS-CoV-2 is greater than previous SARS-CoV, while mechanism of stronger transmission power of SARS-CoV is still unclear. There is no treatment available yet for preventing the pandemic of COVID-19 coronavirus. So, researchers are using phylogenetic, mutational and structural techniques to elucidate the genomic characteristics of SARS-CoV-2 across the world, which will help for vaccine development. Recently, Andersen et al. [7] reported specific mutations in RBD (receptor binding domain) of spike protein of COVID-19 genome. In NCBI database, more than 400 COVID-19 (SARS-CoV-2) assembled genomes are available. Genome sequence analysis can provide plethora of information which can be utilized for vaccine and drug development to cope with COVID-19 pandemic outbreak. In the present study, we retrieved several SARS-CoV-2 genome sequences of Turkey from NCBI and Initiative on Sharing All Influenza Data (GISAID) database up to May 4th, 2020 and were compared with other genomes reported from Saudi Arabia, Iran, America, China, Pakistan, Austria, Denmark, Spain, and Italy to characterize notable genomic features of SARS-CoV-2 and to identify some novel mutations in different regions. Furthermore, transmission and phylogenetic analysis were also conducted to provide significant insight into spread of virus within Turkey.

Materials And Methods

Data Collection

Genomic sequences of SARS-COV-2 from Turkey, Saudi Arabia, Iran, America, China, Pakistan, Denmark, Spain, Italy, China were retrieved using GISAID [8] and NCBI database (http://www.ncbi.nim.nih.gov) (Table 1).

Table 1 List of SARS-Cov-2 genomes used for present analysis

Accession	Location	Collection date
MT327745	Turkey	2020-03-17
MT605818.1	Turkey	2020-04-16
MT675956.1	Turkey	2020-03-27
EPI_ISL_455719	Turkey / Mardin	2020-04-09
EPI_ISL_476832	Turkey / Kocaeli	2020-04-27
EPI_ISL_437334	Turkey / Ankara	2020-03-24
EPI_ISL_437335	Turkey / Denizli	2020-03-25
MT560530.1	Turkey	2020-04-15
MT560525.1	Turkey	2020-04-16
MT560531.1	Turkey	2020-04-16
EPI_ISL_437308	Turkey / Ankara	2020-03-25
EPI_ISL_437315	Turkey / Ankara	2020-03-26
EPI_ISL_437327	Turkey / Agri	2020-03-19
EPI_ISL_428723	Turkey / Aksaray	2020-03-22
EPI_ISL_437329	Turkey / Ankara	2020-03-19
EPI_ISL_437332	Turkey / Istanbul	2020-03-18
EPI_ISL_429861	Turkey / Ankara	2020-03-22
EPI_ISL_429870	Turkey / Sakarya	2020-03-22
EPI_ISL_437309	Turkey / Ankara	2020-03-26
EPI_ISL_428717	Turkey / Kocaeli	2020-03-19
EPI_ISL_428718	Turkey / Kocaeli	2020-03-19
EPI_ISL_428719	Turkey / Siirt	2020-03-21
EPI_ISL_428368	Turkey / Istanbul	2020-04-16
EPI_ISL_437328	Turkey / Tekirdag	2020-03-19
EPI_ISL_437327	Turkey / Agri	2020-03-19
EPI_ISL_437326	Turkey / Istanbul	2020-03-19
EPI_ISL_437326	Turkey / Istanbul	2020-03-19
EPI_ISL_437325	Turkey / Istanbul	2020-03-19
EPI_ISL_437324	Turkey / Istanbul	2020-03-19
EPI_ISL_437323	Turkey / Istanbul	2020-03-19
EPI_ISL_437322	Turkey / Ankara	2020-03-19
EPI_ISL_437314	Turkey / Ankara	2020-03-26
EPI_ISL_437316	Turkey / Denizli	2020-03-25
EPI_ISL_437318	Turkey / Ankara	2020-03-19
EPI_ISL_437319	Turkey / Istanbul	2020-03-19
EPI_ISL_437320	Turkey / Istanbul	2020-03-19
EPI_ISL_437321	Turkey / Istanbul	2020-03-19
EPI_ISL_428722	Turkey / Balikesir	2020-03-22
EPI_ISL_437330	Turkey / Tokat	2020-03-19
EPI_ISL_437306	Turkey / Kocaeli	2020-03-27
EPI_ISL_437307	Turkey / Mardin	2020-03-25
EPI_ISL_437312	Turkey / Kocaeli	2020-03-25
EPI_ISL_437317	Turkey / Ankara	2020-03-27
EPI_ISL_429871	Turkey / Ankara	2020-03-23
EPI_ISL_429872	Turkey / Kocaeli	2020-03-25
EPI_ISL_429873	Turkey / Kocaeli	2020-03-23
EPI_ISL_429868	Turkey / Eskisehir	2020-03-17
EPI_ISL_429867	Turkey / Balikesir	2020-03-17

EPI_ISL_429869	Turkey / Konya	2020-03-17
EPI_ISL_437310	Turkey / Ankara	2020-03-27
EPI_ISL_437311	Turkey / Ankara	2020-03-27
EPI_ISL_437313	Turkey / Kocaeli	2020-03-27
EPI_ISL_429866	Turkey / Afyon	2020-03-16
EPI_ISL_429865	Turkey / Canakkale	2020-03-18
EPI_ISL_429862	Turkey / Ankara	2020-03-22
EPI_ISL_428716	Turkey / Ankara	2020-03-18
EPI_ISL_428715	Turkey / Nevsehir	2020-03-18
EPI_ISL_428714	Turkey / Kastamonu	2020-03-18
EPI_ISL_428713	Turkey / Ankara	2020-03-18
EPI_ISL_428712	Turkey / Karaman	2020-03-17
EPI_ISL_427391	Turkey / Istanbul	2020-04-13
EPI_ISL_424366	Turkey / Kayseri	2020-03-17
EPI_ISL_428720	Turkey / Ankara	2020-03-21
EPI_ISL_437331	Turkey / Ankara	2020-03-25
EPI_ISL_437333	Turkey / Ankara	2020-03-25
EPI_ISL_437304	Turkey / Nevsehir	2020-03-26
EPI_ISL_437305	Turkey / Kocaeli	2020-03-27
EPI_ISL_480261	Turkey / Istanbul	2020-05-03
EPI_ISL_480260	Turkey / Istanbul	2020-05-04
EPI_ISL_480259	Turkey / Istanbul	2020-05-03
EPI_ISL_480258	Turkey / Istanbul	2020-05-03
EPI_ISL_480257	Turkey / Istanbul	2020-05-03
EPI_ISL_480254	Turkey / Istanbul	2020-05-02
EPI_ISL_480253	Turkey / Istanbul	2020-05-01
EPI_ISL_480252	Turkey / Istanbul	2020-05-02
EPI_ISL_480251	Turkey / Istanbul	2020-05-02
EPI_ISL_480250	Turkey / Istanbul	2020-05-02
EPI_ISL_480249	Turkey / Istanbul	2020-04-30
EPI_ISL_480248	Turkey / Istanbul	2020-05-02
EPI_ISL_480244	Turkey / Istanbul	2020-04-23
EPI_ISL_480242	Turkey / Istanbul	2020-04-29
EPI_ISL_480239	Turkey / Istanbul	2020-04-22
EPI_ISL_435144	Spain / Madrid	2020-03-11
EPI_ISL_428689	Spain / Madrid	2020-03-27
EPI_ISL_436397	Spain / Valencia	2020-03-22
EPI_ISL_436341	Spain / Valencia	2020-03-29
EPI_ISL_436377	Spain / Valencia	2020-03-17
MT359866	Spain / Barcelona	2020-03-24
MT233523	Spain / Valencia	2020-03-04
EPI_ISL_417979	Spain / Madrid	2020-03-03
EPI_ISL_417975	Spain / Madrid	2020-03-09
EPI_ISL_417952	Spain / Madrid	2020-03-09
EPI_ISL_428688	Spain / Madrid	2020-03-22
EPI_ISL_419230	Spain / Andalusia	2020-03-05
MT292572.1	Spain / Valencia	2020-03-10
EPI_ISL_428685	Spain / Madrid	2020-03-27
EPI_ISL_436211	Spain / Valencia	2020-03-12
EPI_ISL_436389	Spain / Valencia	2020-03-17
EPI_ISL_412974	Italy / Rome	2020-01-29
EPI_ISL_436718	Italy / Abruzzo	2020-03-19
EPI_ISL_438002	Italy / Cruise ship	2020-03-03

EPI_ISL_438002		
EPI_ISL_437663	Denmark	2020-03-10
EPI_ISL_437668	Denmark	2020-03-27
EPI_ISL_437679	Denmark	2020-03-30
MT320891	Iran	2020-03-09
MT447177	Iran	2020-03-26
EPI_ISL_442523	Iran / Semnan	2020-03-09
EPI_ISL_442044	Iran / Tehran	2020-03-26
EPI_ISL_437762	Saudi Arabia / Makkah	2020-04-02
EPI_ISL_416432	Saudi Arabia / Riyadh	2020-03-07
EPI_ISL_416521	Saudi Arabia	2020-03-10
EPI_ISL_416522	Saudi Arabia	2020-03-10
EPI_ISL_437459	Saudi Arabia / Makkah	2020-03-23
EPI_ISL_437461	Saudi Arabia / Makkah	2020-03-23
EPI_ISL_437463	Saudi Arabia / Makkah	2020-03-23
EPI_ISL_437464	Saudi Arabia / Makkah	2020-03-23
EPI_ISL_437467	Saudi Arabia / Makkah	2020-03-23
EPI_ISL_437475	Saudi Arabia / Jeddah	2020-04-01
EPI_ISL_437484	Saudi Arabia / Madinah	2020-03-29
EPI_ISL_437486	Saudi Arabia / Madinah	2020-03-29
EPI_ISL_437497	Saudi Arabia / Jeddah	2020-04-01
EPI_ISL_437706	Saudi Arabia / Makkah	2020-04-14
EPI_ISL_437710	Saudi Arabia / Makkah	2020-04-14
EPI_ISL_437715	Saudi Arabia / Madinah	2020-04-14
EPI_ISL_437723	Saudi Arabia / Makkah	2020-04-16
EPI_ISL_437729	Saudi Arabia / Makkah	2020-04-16
EPI_ISL_437731	Saudi Arabia / Makkah	2020-04-16
EPI_ISL_437736	Saudi Arabia / Madinah	2020-04-20
EPI_ISL_437739	Saudi Arabia / Madinah	2020-04-20
EPI_ISL_437748	Saudi Arabia / Madinah	2020-04-20
EPI_ISL_402123	China / Hubei / Wuhan	2019-12-24
EPI_ISL_406798	China / Hubei / Wuhan	2019/12/26
EPI_ISL_402132	China / Hubei / Wuhan	2019-12-30
MN908947.1	China / Hubei / Wuhan	2019-12-30
EPI_ISL_403928	China / Hubei / Wuhan	2020-01-01
EPI_ISL_402120	China / Hubei / Wuhan	2020-01-01
EPI_ISL_408515	China / Hubei / Wuhan	2020-01-01
EPI_ISL_414510	China / Shanghai	2020-02-02
EPI_ISL_406801	China / Hubei / Wuhan	2020-01-05
EPI_ISL_412459	China / Hubei / Jingzhou	2020-01-08
EPI_ISL_406030	China / Guangdong / Shenzhen	2020-01-10
EPI_ISL_408486	China / Jiangxi / Pingxiang	2020-01-11
EPI_ISL_408484	China / Sichuan / Chengdu	2020-01-15
EPI_ISL_406594	China / Guangdong / Shenzhen	2020-01-16
EPI_ISL_404227	China / Zhejiang	2020-01-16
EPI_ISL_408481	Asia / China / Chongqing	2020-01-18
EPI_ISL_408482	Asia / China / Shandong / Qingdao	2020-01-19
EPI_ISL_408478	Asia / China / Chongqinq	2020-01-21
EPI_ISL_411060	Asia / China / Fujian	2020-01-21
MN996527.1	Asia / China / Hubei / Wuhan	2019-12-30
EPI_ISL_406223	North America / USA	2020-01-22
EPI_ISL_406036	North America / USA	2020-01-22

EPI_ISL_407214	North America / USA / Washington	2020-01-25
EPI_ISL_407215	North America / USA / Washington	2020-01-25
EPI_ISL_410045	North America / USA / Illinois	2020-01-28
EPI_ISL_408010	North America / USA / California	2020-01-29
EPI_ISL_409067	North America / USA / Massachusetts	2020-01-29
EPI_ISL_408670	North America / USA / Wisconsin	2020-01-31
EPI_ISL_411954	North America / USA / California	2020-02-06
EPI_ISL_411955	North America / USA / California	2020-02-10
EPI_ISL_411956	North America / USA / Texas	2020-02-11
EPI_ISL_413456	North America / USA / Washington	2020-02-20
EPI_ISL_412862	North America / USA / California / Solano	2020-02-23
EPI_ISL_412970	North America / USA / Washington	2020-02-24
EPI_ISL_468163	Pakistan / Islamabad	2020-06-02
EPI_ISL_468162	Pakistan / Islamabad	2020-06-02
EPI_ISL_468161	Pakistan / Islamabad	2020-06-02
EPI_ISL_468160	Pakistan / Islamabad	2020-06-02
EPI_ISL_468159	Pakistan / Islamabad	2020-06-02
MT500122.1	Pakistan / Karachi	2020-03-16
MT262993.1	Pakistan / KPK	2020-03-12
MT240479.1	Pakistan / Gilgit	2020-03-04

Sequence and mutation analysis

The sequences were aligned using Clustal W program. The phylogenetic tree was constructed based on this alignment using the maximum likelihood phylogram in MEGA X software with 1000 replicates [9]. For mutation analysis, sequences of SARS-CoV-2 viral isolates were compared with reference sequence (MN908947.1) from Wuhan.

Secondary structure predictions

The secondary structure of S protein and nonstructural proteins of SARS-COV-2 was predicted by using CFSSP (Chou and Fasman secondary structure prediction), an online server [10].

Results And Discussion

Mutation analysis

Frequent mutations in Turkish isolates

In total, 170 Genome sequences of SARS-COV-2 from different countries were collected from GISAID (https://www.gisaid.org/) and GenBank (http://www.ncbi.nlm.nih.gov/genbank/) sources. In this analysis, 80 SARS- COV-2 genomes sequences submitted from Turkey were investigated to detect different mutations in spike protein (S), nucleocapsid protein (N) and non-structural proteins (Nsp2, Nsp3, Nsp4, Nsp6, NSP12/RdRP) (Table 2).

Table 2 List of mutations identified in 80 SARS-COV-2 genomes of Turkey. The unique mutations in S protein, NSsp2, Nsp3, Nsp4, Nsp12 are marked with *.

List of mutations	Amino acid change	Gene	Type of Mutation	Number of Isolates
23403 A>G	D614G	S	Missense	59
14408 C>T	P4715L	orf1ab	Missense	58
3037C>T		orf1ab	Synonymous	57
241 C>T		5'UTR	Non coding	42
25563G>T	Q57H	ORF3a	Missense	31
18877C>T	L6205L/L280L	orf1ab	Synonymous	29
11083G>T	L3606F	orf1ab	Missense	25
28688T>C		N	Synonymous	23
28881 G>A	R203K	N	Missense	23
28882G>A		N	Synonymous	23
28883 G>C	G204R	N	Missense	23
29742G>T		3'UTR	Non coding	22
1397G>A	V378I	Orf1ab	Missense	21
C26549T		M	Synonymous	13
26735C>T		M	Synonymous	13
884C>A	R207S	Orf1ab/Nsp2	Missense	11*
8653G>T	M2796I	Orf1ab/Nsp4	Missense	11*
19839T>C		Orf1ab	Synonymous	11
7765C>A		Orf1ab/Nsp3	Synonymous	09*
17690C>T	S5809L	Orf1ab	Missense	08
228C>T		5'UTR		08
28854C>T	S194L	N	Missense	08*
22444C>T	D294D	S	Synonymous	08*
2113C>T		Orf1ab/Nsp2	Synonymous	07*
12809C>T	L4182F	orf1ab	Missense	04
10702C>T		Orf1ab	Synonymous	04
26551T>C	V10A	M	Missense	04
22964A>T	I468V	S	Missense	04*
5736 C>T	A1824V	Orf1ab/Nsp3	Missense	03*
13476C>T		orf1ab	Synonymous	03*
3903C>T	P1213L	orf1ab/Nsp3	Missense	03*
881G>A	A206T	orf1ab/Nsp2	Missense	03*
25611C>A		ORF3a	Synonymous	03
8782 C>T		Orf1ab	Synonymous	02*
1059C>T	T265I	orf1ab	Missense	02*
23874C>T	A771V	S	Missense	02*
25275C>T	T1238I	S	Missense	01*
23929C>T	Y789Y	S	Synonymous	01*
22468G>A	T302T	S	Synonymous	01*
25314G>T	G1251V	S	Missense	01*
2416C>T	Y717Y	orf1ab/Nsp2	Synonymous	01*
13730C>T	A4489V	orf1ab	Missense	01
28144T>C	L84S	ORF8	Missense	01
5182T>C		orf1ab	Synonymous	03
27103>T	A194V	M	Missense	01
5477 C>T	H1738Y	orf1ab	Missense	01
6402C>T	P2046L	orf1ab	Missense	01
12741C>A	T4159K	ORF1ab/Nsp8	Missense	01
2997C>T	S911F	orf1ab	Missense	01
13376A>G	T4371A	orf1ab	Missense	01
19484C>T	A6407V	orf1ab	Missense	01
1437C>T	S391F	orf1ab	Missense	01

C7834T		orf1ab	Synonymous	01
944G>A	G227S	orf1ab	Missense	01
26735C>T		M	Synonymous	01
14178C>T		orf1ab	Synonymous	01
25549C>T	L53F	ORF3a	Missense	01
16247C>T	A5328V	orf1ab	Missense	01
6078C>T	A1938V	orf1ab	Missense	01
16616C>T	T5451I	orf1ab	Missense	01
27354A>G		ORF6	Synonymous	01
20464G>A	D6734N	orf1a	Missense	01
9667T>C		orf1ab	Synonymous	01
29197C>T		N	Synonymous	01
7303C>T		orf1ab	Synonymous	01
3464C>T	H1067Y	orf1ab	Missense	01
5192C>T		orf1ab	Synonymous	01
20629C>T	H6789Y	orf1ab	Missense	01
20799T>C		orf1ab	Synonymous	01
28391C>T	R40C	N	Missense	01
25611C>A		ORF3a	Synonymous	03
20668G>A	A6802T	orf1ab	Missense	01
27476C>T	T28I	ORF7a	Missense	01

Based on mutation analysis, 59/80 isolates from turkey have shown a signature 23403A > G (D614G) mutation in the Spike glycoprotein (S) which is clear indicative of a very frequent mutation (73%). Most samples with D614G mutations had high associations with two other mutations (3037 C>T and 14408C > T) in orf1ab region (Table 2). These co-occurring mutations have been recently described as one of major SARS-CoV-2 variant occurring in Europe. For example, Yin [11] detected 15 high frequency SNPs among 558 SARS-COV-2 strains, with most common being 241C > T, 3037C > T, 14408C > T, and 23403A > G in European viral isolates, where the COVID-19 infections by SARS-CoV-2 are generally more severe than other geographical regions.

Regarding the orf1ab region, we also identified previously reported SNPs at the sites 14408 (C>T), 3037 (C>T), 11083 (G>T), 1397 (G>A), 18877 (C>T), 1059 (T>A) and 8782 (C>T). In a report by Banerjee et al. [12], 1134 coding region of orf1ab polyprotein across various states of USA were studied in which four signature mutations T265I (Nsp 2), P4715L (Nsp 12) and P5828L and Y5865C (both at Nsp 13) were identified. In present analysis, 14408C>T (P4715L) and 3037 C>T (F106F) variants of ORF1ab were remarkable to occur at high frequency and presumed to be linked, causing mutations in the RNA dependent RNA polymerase (RdRp/NSP12) site and Nsp3 gene respectively. RdRP/NSP12 is key component of replication/transcription machinery, therefore mutant leucine at 4715 position within RdRP/NSP12 may potentially affect its function, thus increasing the viral mutation rates. Moreover, the proline to leucine was consistently observed as frequent mutation in Europe (51.6%), and North America (58.1%) in previous reports [12, 13]. The variation C3037T was reported to cause synonymous mutation in region encoding Nsp3 and was seen in 57 isolates (71%) from Turkey. Consistent with other studies, the combination of variants at position C3037T, C14408T and A23403G were most common mutations (73%), that exist together in isolates from Turkey.

Other key variants observed in present report includes 25563G>T (Q57H) in ORF3a, along with consecutive series of three variants at position 28881 (G>A), 28882 (G>A) and 28883(G>C) in N proteins. The triple site mutation 28881-28883 that brings change in two amino acid 203-204:RG>KR, is known to play a critical role in virion assembly and structure and had been abundantly seen in US strain previously [14], but also observed recently in Spain, Greece, Vietnam and South America, Mexico, Australia, New Zealand, Belgium, Brazil and Peru [15, 16]. According to GISAID repository, these mutations in N protein were first described in isolates from Netherlands (EPI_ISL_413565) with travel history to Italy, whereas the same mutation was detected in one Italian sample from Abruzzo region (EPI_ISL_436718). In present study, this tri-nucleotide mutation in N protein was observed in eight samples from Turkey with one patient (EPI_ISL_429870) having travel history to Saudi Arabia. This mutation was accompanied by other three mutations such as 241:C>T, 3037:C>T and 14408:C>T. As reported before, the mutated 203/204 region of N protein affects the SR (serine-arginine)-rich motif of the protein (a crucial region for controlling viral transcription and replication) by introducing lysin in between them that might hamper the phosphorylation at serine residue required for normal functioning of N protein [17]. In an investigation by Tylor et al. (2009) [18], significant reduction in pathogenicity has been observed with the deletion of SR domain, therefore it is important to target 203-204:RG>KR positions of the N protein to design the drugs for controlling the disease.

The missense mutation G11083T conferring amino acid change from leucine (L) to phenylalanine (F) in non- structural protein 6 (Nsp6) protein at position 3606 was appeared in 25 samples. In comparison to other countries, it has been observed in Spain, Italy and Iran as well in this study. Previously, it was reported as infrequent mutation from Japan, Netherland and Australia [19,20]. Another substitution (G>A) was found within Nsp2 encoding region of ORF1a at position 1397 and was seen in 26% (21/80) of isolates, which lead to an amino acid change from valine to isoleucine (V378I) having similar isoelectric point. In the study of Pachetti et al. [13], V378I substitution was mainly observed in Oceania viral isolates and less frequently in Asia and North America.

Novel mutations in spike protein

Notably, seven novel mutations were identified in spike (S) protein at different positions which include two synonymous mutations (22444C>T and 23929C>T) and four substitutional mutations (I468V, A771V, T1238I, G1251V) (Table 2). Among synonymous SNPs, C23929T (Y789Y) mutant was observed in only one Turkish sequence (EPI_ISL_455719) of S protein, which was previously reported with high prevalence in Indian strain (39.13%) and one in USA strain (EPI ISL 436898) [12, 21], but not in any other European strain so far. Although, no change in amino acid (Y789Y) at this site, but it was established as a signature SNP in Indian strains, therefore it can be used to trace and monitor the transmission of SARS-CoV-2 within community [21, 22]. The present study also revealed C22444T (S protein) along with C28854T substitution (N protein) as un-common co-evolved mutation both in viral isolates from Turkey (8 samples) and Saudi Arabia (9 samples) (Table S1), which may provide an evidence for travel- associated origin of these mutations to Turkey from Saudi Arabia. However, C22444T mutation has not been

detected in any other European isolate or USA so far but in few Indian strains [12]. On the other hand, C28854T was previously seen in 6/95 samples from different countries [23].

Effect of novel missense mutations in spike protein

D614G characteristic SNP has been considered to play a key role in virus entry and pathogenesis by binding with host ACE2 receptor via variable amino acid residues from 331 to 524 of S protein called receptor-binding domain (RBD) in S1 subunit [24-25]. Strikingly, a substitutional mutation (I468V) at 468th position was found in four Turkish isolates (EPI_ISL_437316, EPI_ISL_437322, EPI_ISL_437323 and EPI_ISL_437324) in RBD of the spike protein which was not previously reported in strains of any other country. At this site, isoleucine (I) was replaced by valine (V) at position 468. Since both amino acids are hydrophobic in nature having C beta branched residues, thus any change due to this mutation did not cause any functional change in protein, as predicted by our secondary structure analysis (Fig. 1A). Despite these, this site is more prone to mutate further that may change the binding property of S protein to ACE2 receptor. However, the impact of such mutation is needed to explore further to unrevealed the crucial role of this binding site for interaction with host ACE2. Similarly, two strains (EPI_ISL_437314, EPI_ISL_437316) showed a missense mutation at position 771 exchanging alanine (A) with valine (V) in S protein. From the results of secondary structure data, it was demonstrated that helix structure has been replaced at positions 770 and 771 due to addition of two sheets at these sites (Fig. 1B). Alanine to valine substitution as residue 771 was previously seen in one Belgium strain [21].

Additionally, threonine (polar amino acid) to Isoleucine (non-polar amnio acid) substitution due to T1238I mutation and change of glycine to bulkier valine at position 1251 (G1251V) were specifically found in EPI_ISL_429870 and EPI_ISL_437317 strains respectively. The detailed secondary structure analysis revealed that change of threonine to Isoleucine makes it hydrophobic and induce structural alteration in that domain with an addition of helix at position 1238 (Fig. 1C), while G1251V mutant have attained changes in secondary structure at mutation locus with six additional sheets from position 1448 to 1253 and a helix at site 1254 (Fig. 1D). Some of earlier reports suggests that substitutional mutations like threonine to Isoleucine and alanine to valine in EBOV glycoprotein (GP) has increased the viral infectivity of Ebola virus in humans [26, 27], therefore, it is assumed that mutant residues at 771, 1238 and 1254 positions in spike protein may lead to an alteration in the way the spike interacts with the receptor, changing the infectivity as these mutations lies in S2 subunit of S protein, but this require further research and more sequencing.

Novel mutations in Nsp2

In total five uncommon mutations within the Nsp2 region (C1059T, C2416T, G881A, C884T, C2113T) were identified in this report (Table 2). Some of these infrequent mutations (C8782T and T1059A) were appeared to be predominate in other countries [28, 29]. Nsp2 is postulated to have an important role in the host cell survival pathway via interaction with prohibitin (PHB) and prohibitin 2(PHB2) [30]. A change of nonpolar amino acid (Alanine) to a polar amino acid (Threonine) at position 206 (G881A mutation) within Nsp2 gene was uniquely appeared in three sequences (EPI_ISL_480239, EPI_ISL_428723,

EPI_ISL_429863) from Turkey only. Similarly, 884C>T mutation that resulted in the change of arginine codon (R) to cysteine (C) at position 207 were displayed by 11 Turkish viral isolates and one Pakistani strain (MT240479.1) (Table S1). Furthermore, two samples of Turkey (EPI_ISL_437309; EPI_ISL_437315), along with one strain from Spain (EPI_ISL_428688) and Denmark (EPI_ISL_437668) showed substitutional change (Threonine to Isoleucine) at site 265 (C1059T mutation). In addition, a synonymous mutation (2416C>T) was reported in two infected individuals (EPI_ISL_428712; EPI_ISL_437332) from turkey with travel history from Iran and in three samples (EPI_ISL_468162; EPI_ISL_468161; EPI_ISL_468160) from Pakistan. Previously, T265I mutation was detected exclusively in American population at frequency of 43 % making it a signature SNP for USA, whereas it was found in very low frequency in Asia (4.8%) [30]. Likewise, a change in nucleotide (C>T) was observed at position 2113 in seven Turkish samples and two isolates of Saudi Arabia, but none from any other sequence in this report.

Effect of missense mutations on Nsp2

Our secondary structure prediction analysis showed that A206T, R207C and T265I mutations has caused structural alterations within Nsp2 domain (Fig. 1E). For example, a loss in α-helix was noted at positions 203, 204 and 206 with addition of two sheets at 203 and 204 due to mutation of A206T. Similarly, R207C mutation resulted in substitution of α-helix with a sheet structure at position 204 and an additional turn at 210 position (Fig. 1F). Further, there is an addition of sheet structure at position 266 when threonine is substituted by isoleucine in mutant T265I (Fig. 1G). Since all these mutations were identified in Nsp2 domain that may cause structural alterations, therefore it is essential to consider these mutational spectra while designing new antiviral therapeutics targeting viral orf1ab. However, further experimental work is required to study the effect of these mutations on SRAS-CoV-2.

Novel mutations in Nsp3

In SARS-CoV virus, NSP3 has been proposed to work with NSP4 and NSP6 to induce double membrane vesicles (DMV), that serve as important unit for replication/transcription complex [31]. In this report, some novel variants in Nsp3 region were also detected at the sites 3903 (C>T), 5736 (C>T) and 7765 (C>A). Out of these three mutations, two were missense mutations (C3903Tand C5736T) that brings amino acid changes from proline to leucine (P1213L) and alanine to valine (A1824V) respectively and each were exclusively seen in isolates from Turkey only, while C7765A was a silent mutation and was observed in 11 % (9/80) of Turkish samples, while only one viral isolate from Saudi Arabia (EPI_ISL_437463) had this variation (Table S1).

Effect of missense mutations on Nsp3

In mutant A1824V, there is loss of turn and addition of a sheet structure at position 1825 as shown by secondary structure prediction (Fig. 1H), that might result in significant functional implications, whereas, no substantial change in secondary structure was observed for mutant P1213L (Fig. 1I). It is important to

evaluate these mutations in Nsp3, as this gene has been reported to harbour many mutations that resulted in the evolution of beta coronaviruses with extensive selection pressure [32].

Novel mutations in Nsp4 and Nsp12

Apart from frequent mutations, two mutations in NSP4 (C8782T and G8653T) and one in NSP12/RdRP gene (13730C>T) were found completely unique to Turkey. A silent mutation at site 8782 (C>T) within Nsp4 was displayed by two infected individuals (EPI_ISL_428718 and EPI_ISL_437317) from Turkey and seven samples from Spain, describing it as infrequent mutation, however, it was present at high frequency in both Oceania and North America isolates in previous reports [13]. G8653T is another Nsp4 mutation (missense) which substitute methionine (M) into isoleucine (I) at site 2796 and was found in 11 Turkish strains, not observed in any other viral isolates from Europe so far. However, this mutation was detected by Joshi and Paul [33] in nine Indian samples and two isolates from Kuwait. At NSP12/RdRP gene, alanine/valine substitution was observed at site 4489 in a single sequence (EPI_ISL_455719, Turkey / Mardin) on April 9th, 2020 which was not found in any other sequence in this report. In a previous report by Maitra et al. [34], this unique mutation was also depicted in two infected individuals of India.

Effect of missense mutations on Nsp4 and Nsp12/RdRp

Further, our secondary structure prediction analysis showed that mutant M2796I causes changes in secondary structure within Nsp4 as there is replacement of α-helix with sheet structure at position 2795 and addition of turn at 2798 site (Fig. 1J), which might affect the interaction between Nsp3 and Nsp4 and viral replication in SARS-CoV-2. Normally, valine side chain is larger than alanine, and substitution of valine in Nsp12 resulted in the of loss of α-helixes at position 4486,4487, and 4488 with the addition of five sheets from positions 4486 to 4490 as revealed by our secondary structure predications (Fig. 1K), therefore, this substitution might have functional consequences that can potentially affect the viral replication and mutagenic capabilities of SARS-CoV-2.

Importantly, it will be quite interesting to validate the effects of these substitutions on non-structural proteins as one of previous study proposed that alanine/valine substitution in NS2A (non-structural protein) in Zika virus impairs viral RNA synthesis and results in viral attenuation (Márquez -Jurado et al., 2018) [35]. Similarly, valine substitution in RdRp protein in Indian SARS-CoV-2 isolates has caused structural alteration that impairs packing of the protein (Chand et al. 2020) [36]. Thus, functional characterization of these mutations investigated in our study needs to be carried out to understand the exact role of these mutations and to develop strategies for vaccine designing that target this virus.

Transmission and phylogenetic analysis

Interestingly, 13 viral isolates of turkey harboring 23403 A>G, 3037 C>T, and 14408 C>T concurrent mutations having travel history from Saudi Arabia (Table 3). Of these, 6 cases were reported from Ankara city while remaining were from Aksaray, Sakarya, Tekirdag, Kocaeli, Kastamonu, Konya and Afyon. Similarly, 6 samples had association with Iran travel history and maximum cases (5/6) were noted from

Istanbul mentioned in GISAID. Among Iran-travel linked isolates, 4 samples (EPI_ISL_437319, EPI_ISL_437324, EPI_ISL_437325,EPI_ISL_437327) are defined by the presence of three co-mutations (G1397A, T28688C and G29742T, a Europe based introduction), while remaining 2 samples (EPI_ISL_437326, EPI_ISL_437332) lacked these mutations (Table 3). However, G1397A and T28688C substitution was also observed in one patient who had travelled to Taiwan, this probably suggesting the spread of COVID-19 infection to Turkey from multiple countries especially from Saudi Arabia and Iran. In a similar analysis, Eden et al. [22] provided an evidence for travel associated SARS-CoV-2 origin to Australia from Iran. Our findings may also contribute to better understand the diversity of circulating SARS-COV-2 and origin of imported cases to Turkey.

Table 3 List of Turkish strains with travel history.

Accession	Virus name	Location	Collection date	Travelling history
EPI_ISL_429866	hCoV-19/Turkey/HSGM- 4236/2020	Turkey / Afyon/west	2020-03-16	Saudi Arabia
EPI_ISL_429869	hCoV-19/Turkey/HSGM- 4701/2020	Turkey / Konya	2020-03-17	Saudi Arabia
EPI_ISL_428716	hCoV-19/Turkey/HSGM- 5711/2020	Turkey / Ankara	2020-03-18	Saudi Arabia
EPI_ISL_428714	hCoV-19/Turkey/HSGM- 5602/2020	Turkey / Kastamonu	2020-03-18	Saudi Arabia
EPI_ISL_429862	hCoV-19/Turkey/HSGM- 8970/2020	Turkey / Ankara	2020-03-22	Saudi Arabia
EPI_ISL_437332	hCoV-19/Turkey/HSGM- 4698/2020	Turkey / Istanbul	2020-03-18	Iran
EPI_ISL_428718	hCoV-19/Turkey/HSGM- 5770/2020	Turkey / Kocaeli	2020-03-19	Saudi Arabia
EPI_ISL_437328	hCoV-19/Turkey/HSGM- 7668/2020	Turkey / Tekirdag	2020-03-19	Saudi Arabia
EPI_ISL_437326	hCoV-19/Turkey/HSGM- 1432/2020	Turkey / Istanbul	2020-03-19	Iran
EPI_ISL_437325	hCoV-19/Turkey/HSGM- 1495/2020	Turkey / Istanbul	2020-03-19	Iran
EPI_ISL_437324	hCoV-19/Turkey/HSGM- 1476/2020	Turkey / Istanbul	2020-03-19	Iran
EPI_ISL_437327	hCoV-19/Turkey/HSGM- 1458/2020	Turkey / Agri	2020-03-19	Iran
EPI_ISL_437323	hCoV-19/Turkey/HSGM- 1481/2020	Turkey / Istanbul	2020-03-19	Taiwan
EPI_ISL_437319	hCoV-19/Turkey/HSGM- 1490/2020	Turkey / Istanbul	2020-03-19	Iran
EPI_ISL_429870	hCoV-19/Turkey/HSGM- 8990/2020	Turkey / Sakarya	2020-03-22	Saudi Arabia
EPI_ISL_428723	hCoV-19/Turkey/HSGM- 8964/2020	Turkey / Aksaray	2020-03-22	Saudi Arabia
EPI_ISL_429871	hCoV-19/Turkey/HSGM- 10241/2020	Turkey / Ankara	2020-03-23	Saudi Arabia
EPI_ISL_437317	hCoV-19/Turkey/HSGM- 1027/2020	Turkey / Ankara	2020-03-27	Saudi Arabia
EPI_ISL_437329	hCoV-19/Turkey/HSGM- 6204/2020	Turkey / Ankara	2020-03-19	Saudi Arabia
EPI_ISL_437331	hCoV-19/Turkey/HSGM- 1014/2020	Turkey / Ankara	2020-03-25	Saudi Arabia

Phylogenetic analysis of SARS-COV-2 genomes suggests that Turkish strain shares close relationships with isolates from Saudi Arabia predicting the possibility of common origin (Fig. 1 A and B). However, the viral genomes of Turkey were dispersed throughout the phylogenetic tree indicating multiple independent introductions to the country. In phylogram two distinct clades were categorized as Cluster 1 and Cluster 2. The first cluster 1 had a key mutation at 1397G>A, 11083G>T, 28688T>C, and 29742G>T, while second cluster had 23403 A>G, 3037 C>T, and 14408 C>T dominant mutations. Moreover, majority of viral isolates in Turkey showed L type characteristics and form monophyletic clade, while S type were present in limited numbers. SNP with “T” at site 28144 encoding leucine were classified as L type, while “C” at this position encoding serine is referred as S type. Comparatively, L type was speculated more aggressive and contagious than S type [12]. Thus, it appears that L strain is predominant strain circulating among Turkish population. In addition, samples having travel connection to Saudi Arabia and Iran were showing monophyletic origin within their respective clusters in phylogenetic tree.

In conclusion, the functional characterization of novel mutations investigated in our study needs to be carried out to understand the exact role of these variations. Furthermore, the above-mentioned mutations might pave way towards the identification of less virulent strains and development of vaccines for large repertoire of strains. Phylogenetic and transmission analysis revealed that spread of SARS-COV-2 to Turkey was due to multiple independent sources of introductions and viral isolates of Saudi Arabia and Turkey are closely related with one another.

Declarations

Competing Interests

The authors declare that they have no competing interests.

Author contributions

All authors contributed to the development, analysis, and drafting of this article.

Funding

No funding was received for this research.

References

1. Gorbalenya A et al. (2020) The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology 5: 536–544.
2. Transmission of SARS-CoV-2: implications for infection prevention precautions (2020). https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2 implications-for-infection-prevention-precautions. Accessed 5th May, 2020

3. Chen J (2020) Pathogenicity and transmissibility of 2019-nCoV—a quick overview and comparison with other emerging viruses. Microbes and Infection 22: 69–71.
4. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell host & microbe. 27(3): 325-328
5. Hoffmann M, Kleine-Weber H, Krüger N, Müller M, Drosten C, Pöhlmann S. The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. bioRxiv. 2020:2020.01.31.92904.
6. Guan W.J. Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei Cl, Hui DS et al. (2020) Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 382:1708-1720
7. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. Nat Med 100:163.
8. Shu Y, McCauley J (2017) GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill 22(13):30494. doi:10.2807/1560- ES.2017.22.13.30494
9. Kumar S, Stecher G, Tamura K. MEGA7 (2016) Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870-4.
10. Ashok Kumar T (2013) CFSSP: Chou and Fasman secondary structure prediction server. Wide Spectrum 1:15–19 DOI 5281/zenodo.50733.
11. Yin C (2020) Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics 19: 1–12.
12. Banerjee S, Seal S, Dey R, Bhattacharjee P (2020) Mutational spectra of SARS-CoV-2 orf1ab polyprotein and Signature mutations in the United States of America.
13. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E et al (2020) Emerging SARS-CoV 2 mutation hot spots include a novel RNA - dependent - RNA polymerase variant. Journal of Translational Medicine 18 (179): 1-9. doi: 10.1186/s12967- 020-02344-6
14. Ayub MI (2020) Reporting Two SARS-CoV-2 Strains Based on A Unique Trinucleotide-Bloc Mutation and 351 Their Potential Pathogenic Difference. Preprints 2020040337. doi:10.20944/preprints202004.0337.v1
15. Laamarti M, Alouane T, Kartti S, Chemao-Elfihri MW, Hakmi M, Essabbar A, Laamart M, Hlali H, Allam L et al (2020) Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geodistribution and a rich genetic variations of hotspots mutations. bioRxiv. https://doi.org/10.1101/2020.05.03.074567
16. Lorusso A, Calistri P, Mercante MT, Monaco F, Portanti O, Marcacci M, Cammà C, Rinaldi A, Mangone I, Di Pasquale A, Iommarini MA (2020) “One-Health” approach for diagnosis and molecular characterization of SARS-CoV-2 in Italy. One Apr 19:100135.
17. Peng, TY, Lee KR, Tarn WY (2008) Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. FEBS J 275: 4152–4163

18. Tylor S et (2009) The SR-rich motif in SARS-CoV nucleocapsid protein is important for virus replication. Can. J. Microbiol. 55: 254–260.
19. Mercatelli D, Giorgi FM (2020) Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Preprints. 338 2020; 2020040529. doi: 20944/preprints202004.0529.v1
20. Benvenuto D, Angeletti S, Giovanetti M, et (2020) Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy [published online ahead of print]. J Infect 2020;S0163-4453(20)30186-9. doi:10.1016/j.jinf.2020.03.058.
21. Samyuktha V, Kumar V N (2020) Emergence of RBD and D614G Mutations in Spike Protein: An Insight from Indian SARSCoV-2 Genome Analysis. doi: 20944/preprints202006.0032.v1
22. Eden JS, Rockett R, Carter I, Rahman H, De Ligt J et al (2020) An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. Virus Evol 6 (1): 1-4. doi: 1093/ve/veaa027
23. Wang C, Liu Z, Chen Z, Huang X, Xu M et al. (2020a) The establishment of reference sequence for SARS- CoV-2 and variation analysis. J Med Virol 92 (6): 667-674. doi: 1002/jmv.25762
24. Tai W, He L, Zhang X, Pu J, et al. (2020) Characterization of The Receptor-Binding Domain (RBD) Of 2019Novel Coronavirus: Implication for Development Of RBD Protein As A Viral Attachment Inhibitor And Vaccine.
25. Chen WH, Hotez PJ, Bottazzi ME (2020) Potential for developing a SARS-CoV receptor-binding domain (RBD) recombinant protein as a heterologous human vaccine against coronavirus infectious disease (COVID)-19. Hum Vaccin Immunother 1-4.
26. Ueda MT, Kurosaki Y, Izumi T, Nakano Y, Oloniniyi OK et al. (2017) Functional mutations in spike glycoprotein of Zaire ebolavirus associated with an increase in infection efficiency. Genes Cells 22: 148–
27. Kurosaki Y, Ueda MT, Nakano Y, Yasuda J, Koyanagi Y, Sato K, Nakagawa S (2018) Different effects of two mutations on the infectivity of Ebola virus glycoprotein in nine mammalian species. J Gen Virol 99: 181–186.
28. Wang JT, Lin YY, Chang SY, et al. (2020b) The role of phylogenetic analysis in clarifying the infection source of 340 a COVID-19 patient. J Infect. 2020;S0163-4453(20)30159-6. doi:10.1016/j.jinf.2020.03.031
29. Phelan J, Deelder W, Ward D, Campino S, Hibberd ML, Clark TG (2020) Controlling the SARS-CoV-2 outbreak, insights from large scale whole genome sequences generated across the world. bioRxiv. 2020;2020.04.28.066977. https://doi.org/10.1101/2020.04.28.066977.
30. Rouchka EC, Chariker JH, Chung D (2020) Phylogenetic and Variant Analysis of 1,040 SARS-CoV-2 Genomes. Preprints. doi: 20944/preprints202005.0396.v1.
31. Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W (2017) Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology 510:165–174
32. Wu C, Liu Y, Yang Y, et al. (2020) Analysis of Therapeutic Targets for SARS-CoV-2 and Discovery of Potential Drugs by Computational Methods. Acta Pharmaceutica Sinica B,

33. Joshi A, Paul S (2020) Phylogenetic Analysis of the Novel Coronavirus Reveals Important Variants in Indian Strains. Preprint, doi: https://doi.org/10.1101/2020.04.14.041301
34. Maitra A1, Sarkar MC, Raheja H, et al. (2020) Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. J Biosciences. DOI: 1007/s12038-020-00046-1
35. Márquez-Jurado S, Nogales A, Ávila-Pérez G, Iborra FJ, Martínez-Sobrido L, Almazán F (2018) An alanine- to-valine substitution in the residue 175 of Zika virus NS2A protein affects viral RNA synthesis and attenuates the virus in vivo. Viruses 10(10):547.
36. Chand GB, Banerjee A, Azad GK (2020) Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. bioRxiv. 2020 Jan

Figures

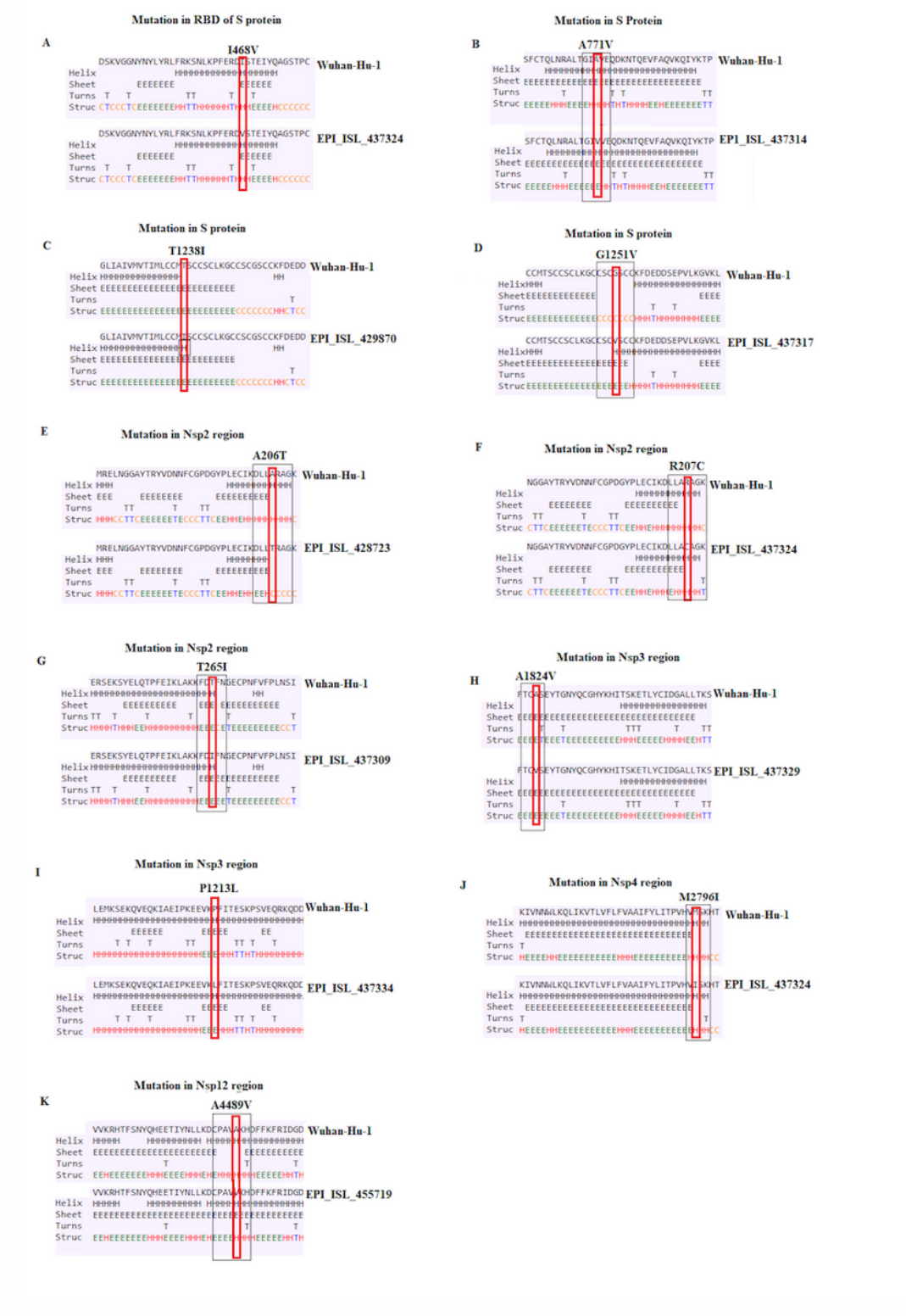

Figure 1

Prediction of secondary structure of S protein, Nsp2, Nsp3, Nsp4, Nsp12 regions. (A-D) demonstrate mutations in S protein; (E-G) represent mutations in Nsp2 region; (H-I) represent mutations in Nsp3 region; (J) represent mutation in Nsp4 region; (K) represent mutation in Nsp12 region. The small rectangular box shows mutant residues (highlighted in red box). The difference of secondary structure between Wuhan and Turkish isolates are highlighted with position of black box in respective panels.

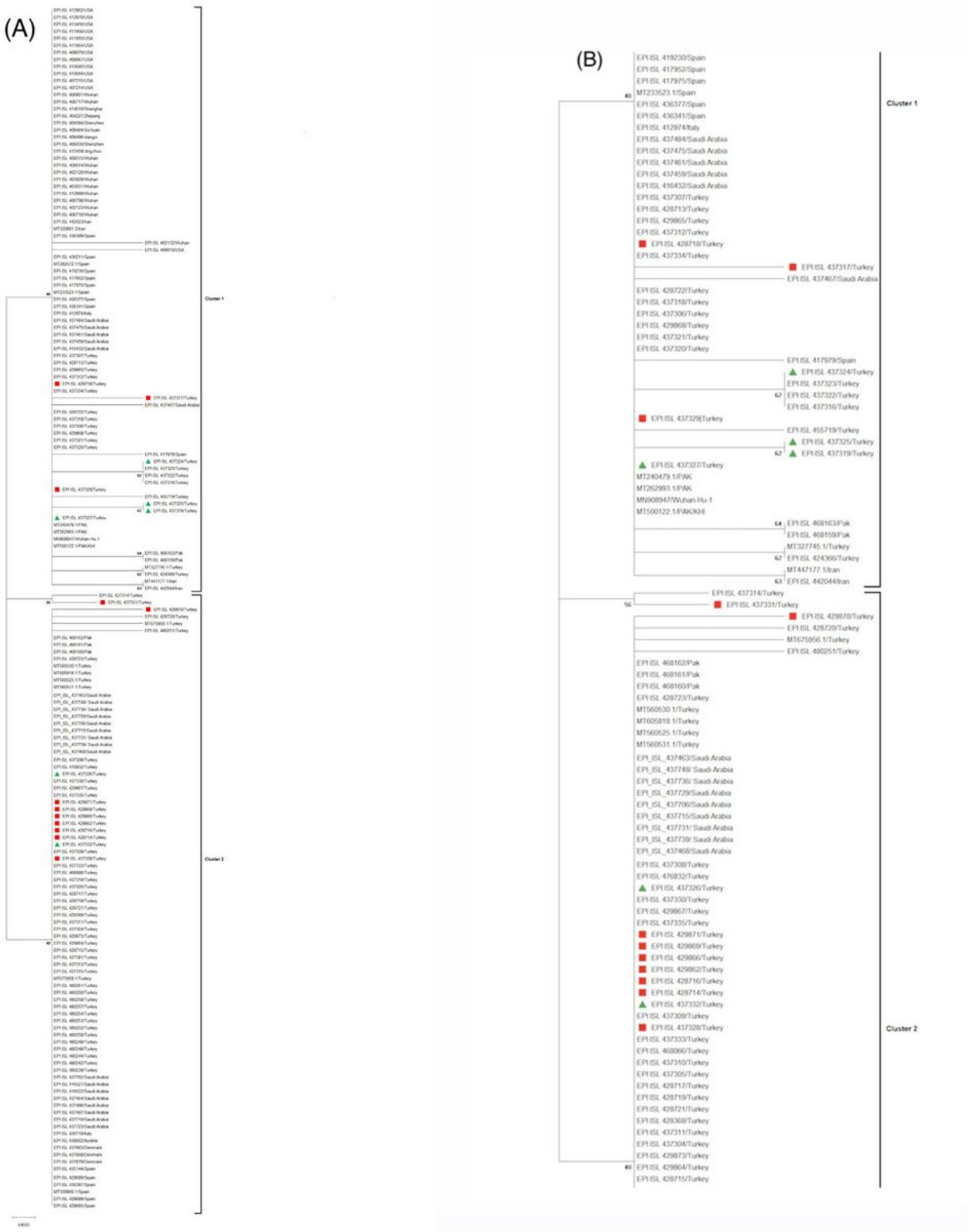

Figure 2

(A) Phylogenetic analysis of SARS-CoV-2 genome sequences highlighting a clade of imported cases from Saudi Arabia and Iran to Turkey. (B) Sub-tree showing the informative branch containing imported cases to Turkey from Saudi Arabia highlighted with red squares) and Iran (highlighted with green triangles).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Rehmanetal.supplementarytable.docx