





Mutation Profile of SARS-CoV-2 Genome Sequences Originating from Eight Israeli Patient Isolates

 Galia Zaide,^a Inbar Cohen-Gihon,^a Ofir Israeli,^a Dana Stein,^a Ohad Shifman,^a Shay Weiss,^b Irit Simon,^c Orly Laskar,^b Adi Beth-Din,^a  Anat Zvi^a

^aDepartment of Biochemistry and Molecular Genetics, Israel Institute for Biological Research, Ness-Ziona, Israel

^bDepartment of Infectious Diseases, Israel Institute for Biological Research, Ness-Ziona, Israel

^cDepartment of Biotechnology, Israel Institute for Biological Research, Ness-Ziona, Israel

ABSTRACT We report the genome sequences and the identification of genetic variations in eight clinical samples of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Samples were collected from nasopharyngeal swabs of symptomatic and asymptomatic individuals from five care homes for elderly and infirm persons in Israel. The sequences obtained are valuable, as they carry a newly reported nonsynonymous substitution located within the nucleoprotein open reading frame.

Shortly after a severe acute respiratory syndrome emerged in Wuhan, China, in December 2019 (1, 2), a new *Betacoronavirus* strain of the *Coronaviridae* family named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified as the etiological agent of a disease that was later termed coronavirus disease 19 (COVID-19) (2, 3). In this report, we describe the sequencing of eight SARS-CoV-2 samples obtained from specimens from five care homes for elderly and infirm persons in Israel. This study is in line with the ethical statement of the associate director general of the Israeli Ministry of Health. The individuals were initially identified as positive for COVID-19 by reverse transcriptase quantitative PCR (RT-PCR) and exhibited low cycle threshold (C_T) values ranging from 12.8 to 16.8, implying a high viral load. Partial clinical information indicated that at least 2 of the 8 samples (i.e., EPI_ISL_594157 and EPI_ISL_594158) originated from asymptomatic individuals.

Samples were collected directly from swabs, and RNA was extracted with a QIAamp viral RNA minikit (Qiagen) according to the manufacturer's protocol, using 60 μ l of AVE buffer for elution. A SMARTer stranded total RNA-Seq pico input mammalian v2 kit (TaKaRa) was used for library construction prior to sequencing on a MiSeq instrument (Illumina). Whole-genome, paired-end sequencing was conducted in a duplex or triplex format with a read length of 150 nucleotides.

FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>) with default settings was used for quality control of the data. Trimming and removal of low-quality

TABLE 1 Genome features of eight SARS-CoV-2 clinical samples

Sample	Genbank accession no.	Total no. of reads	No. of mapped reads	Avg coverage (×)	Assembly length (bp)	Overall G+C content (%)
NH-MA	MW227568	1,151,461	6,406	37	29,894	37.94
NH-GD3	MW201578	2,299,402	20,459	128	29,895	37.94
NH-NM	MW193889	1,759,712	51,125	214	29,899	37.94
NH-GD2	MW201577	1,245,093	28,057	143	29,870	37.94
NH-GD1	MW237708	1,099,546	14,081	62	29,895	37.94
NH-AS	MW201576	3,106,246	36,205	245	29,927	37.92
NH-M2	MW194121	6,485,364	59,944	297	29,930	37.91
NH-M1	MW228070	3,982,658	16,558	113	29,942	37.99

Citation Zaide G, Cohen-Gihon I, Israeli O, Stein D, Shifman O, Weiss S, Simon I, Laskar O, Beth-Din A, Zvi A. 2021. Mutation profile of SARS-CoV-2 genome sequences originating from eight Israeli patient isolates. *Microbiol Resour Announc* 10:e01387-20. <https://doi.org/10.1128/MRA.01387-20>.

Editor John J. Dennehy, Queens College

Copyright © 2021 Zaide et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Galia Zaide, galiaz@iibr.gov.il.

Received 3 December 2020

Accepted 10 December 2020

Published 7 January 2021

Nucleotide position	Reference base	Mutant base	Substitution	Mutation found in samples								Gene	Type of mutation
				NH-MA	NH-GD3	NH-NM	NH-GD2	NH-GD1	NH-AS	NH-M2	NH-M1		
241	C	T	-	T	T	T	T	T	T	T	T	Intergenic	Noncoding
313	C	T	L16L								T	orf1ab	Synonymous
541	C	T	L92L								T	orf1ab	Synonymous
683	C	T	L140L	T						T		orf1ab	Synonymous
1059	C	T	T85I	T	T	T	T	T	T			orf1ab	non-synonymous
1213	C	T	C316C			T						orf1ab	Synonymous
1528	C	T	A421A							T	T	orf1ab	Synonymous
2836	C	T	C857C							T	T	orf1ab	Synonymous
3037	C	T	F924F	T	T	T	T	T	T	T		orf1ab	Synonymous
4510	G	A	V1415V							A	A	orf1ab	Synonymous
6402	C	T	P1228L	T						T		orf1ab	non-synonymous
6807	C	T	T1363		T		T					orf1ab	non-synonymous
8905	C	T	D2880D							T	T	orf1ab	Synonymous
10249	T	C	N3328N	C						C		orf1ab	Synonymous
11083	G	T	L37F							T		orf1ab	non-synonymous
11200	C	T	A3645A			T						orf1ab	Synonymous
11916	C	T	S25L		T	T	T	T				orf1ab	non-synonymous
14408	C	T	P323L	T	T	T	T	T	T	T	T	orf1ab	non-synonymous
15438	G	T	M666I							T	T	orf1ab	non-synonymous
16935	G	T	M233I	T								orf1ab	non-synonymous
17122	G	T	A296S		T							orf1ab	non-synonymous
17427	G	T	V1250V	T						T		orf1ab	Synonymous
18105	G	T	Q1476H			T						orf1ab	non-synonymous
18264	C	A	T1529T			A						orf1ab	Synonymous
18998	C	T	A320V		T	T	T	T				orf1ab	non-synonymous
20628	C	T	G2317G				T					orf1ab	Synonymous
20646	C	T	Y2323Y							T		orf1ab	Synonymous
21193	G	T	D179Y	T								orf1ab	non-synonymous
21575	C	T	L5F	T								Spike	non-synonymous
21761	G	T	A67S	T								Spike	non-synonymous
22347	C	T	A262V	T								Spike	non-synonymous
23125	A	T	P521P								T	Spike	Synonymous
23403	A	G	D614G	G	G	G	G	G	G	G	G	Spike	non-synonymous
25563	G	T	Q57H		T	T	T	T	T			ORF3a	non-synonymous
25785	G	T	W131C								T	ORF3a	non-synonymous
25970	G	T	W193L							T		ORF3a	non-synonymous
25971	G	A	W193L							A		ORF3a	non-synonymous
26060	C	T	T223I	T								ORF3a	non-synonymous
26217	G	T	L275F		T	T	T	T				ORF3a	non-synonymous
26709	G	T	A63S		T		T	T				M	non-synonymous
27925	C	T	T11I		T		T	T				ORF8	non-synonymous
28320	C	T	T16M	T								N	non-synonymous
28421	G	T	A50S			T						N	non-synonymous
28851	G	T	S193I								T	N	non-synonymous
28854	C	T	S194L	T						T		N	non-synonymous
28881	G	A	R203K							A	A	N	non-synonymous
28882	G	A	R203R							A		N	Synonymous
28883	G	C	G204R							C	C	N	non-synonymous
28932	C	T	A220V			T						N	non-synonymous
29518	C	T	D415D	T								N	Synonymous
29540	G	A	-		A	A	A	A				Intergenic	Noncoding
29752	A	T	-	T	T	T	T	T	T	T	T	Intergenic	Noncoding

FIG 1 Mutations detected in the SARS-CoV-2 sequences generated in this study. The reference base was retrieved from the reference Wuhan strain (GenBank accession number [NC_045512](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512)). Samples are named according to the isolate feature in the corresponding GenBank record (see Table 1).

reads were performed using Trim Galore! v0.6.3 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default settings. Bowtie 2 (4) with default parameters was used for filtering of the results and for mapping the filtered reads against the reference Wuhan strain (GenBank accession number [NC_045512](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512)). Reads mapped to SARS-CoV-2 were used as input data for the SPAdes assembler v3.13.0 (5) or the DNASTar software (SeqMan NGen v17.0; DNASTar, Madison, WI), resulting in a single contig for each sample. The genomic features of the samples are summarized in Table 1. Variant calling was performed using the SAMtools software package (6) with default parameters; a variant quality score cutoff of 100 was applied for all samples. A phylogenetic analysis generated using Nextstrain (7), rooted relative to the early samples from Wuhan, revealed that two of the eight samples (i.e., EPI_ISL_594155 and EPI_ISL_594156) belong to clade 20B, while the rest belong to clade 20C.

The variant calling process revealed a total of 52 unique single-nucleotide polymorphism (SNP) replacements. A total of 31 substitutions were nonsynonymous, 4 of which mapped to the Spike coding region; 18 substitutions were of the synonymous type, and the remaining 3 substitutions occurred in noncoding regions (Fig. 1). The eight samples share one common mutation in an intergenic region (position 241, C to T) and two common mutations in coding regions (positions 23403, A to G, and 14408, C to T), resulting in the well-documented D614G substitution and the P323L replacement, respectively (Fig. 1). Apart from the abundant D614G replacement, six other nonsynonymous abundant replacements found in this study (i.e., T85I, L37F, S25L, P323L, A320V, and Q57H; Fig. 1) were previously reported as a result of hot spot mutations (8–10).

While most of the nonsynonymous replacements were previously reported (11), the A50S substitution (located in the nucleocapsid protein) identified in the EPI_ISL_594161 sample, was not documented before (GISAID [12, 13], as of November 2020).

Although several papers documented a list of viral factors that are correlated with COVID-19 severity (9, 14–16), there is still more to it than meets the eye. Thus, mapping and identification of new mutations may contribute to a better understanding of the viral factors related to clinical outcomes of the disease.

Data availability. The genome sequences have been deposited at the GISAID EpiCoV coronavirus SARS-CoV-2 platform database under the identifiers EPI_ISL_594155, EPI_ISL_594156, EPI_ISL_594157, EPI_ISL_594158, EPI_ISL_594159, EPI_ISL_594160, EPI_ISL_594161, and EPI_ISL_594162 and in the NCBI GenBank database under the accession numbers [MW228070](https://www.ncbi.nlm.nih.gov/nuccore/MW228070), [MW194121](https://www.ncbi.nlm.nih.gov/nuccore/MW194121), [MW201576](https://www.ncbi.nlm.nih.gov/nuccore/MW201576), [MW227568](https://www.ncbi.nlm.nih.gov/nuccore/MW227568), [MW237708](https://www.ncbi.nlm.nih.gov/nuccore/MW237708), [MW201577](https://www.ncbi.nlm.nih.gov/nuccore/MW201577), [MW193889](https://www.ncbi.nlm.nih.gov/nuccore/MW193889), and [MW201578](https://www.ncbi.nlm.nih.gov/nuccore/MW201578). The raw reads have been submitted to the NCBI Sequence Read Archive under the study reference number [PRJNA672811](https://www.ncbi.nlm.nih.gov/sra/PRJNA672811).

ACKNOWLEDGMENT

We thank Emanuelle Mamroud for fruitful discussions and support throughout the project.

REFERENCES

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395:497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus Investigation, Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>.
- Alouane T, Laamarti M, Essabbar A, Hakmi M, Bouricha EM, Chemao-Elfihri MW, Kartti S, Boumajdi N, Bendani H, Laamrti R, Ghriif F, Allam L, Aanniz T,

- Ouadghiri M, El Hafidi N, El Jaoudi R, Benrahma H, Elattar J, Mentag R, Sbabou L, Nejjari C, Amzazi S, Belyamani L, Ibrahim A. 2020. Genomic diversity and hotspot mutations in 30,983 SARS-CoV-2 genomes: moving toward a universal vaccine for the “confined virus”? bioRxiv <https://www.biorxiv.org/content/10.1101/2020.06.20.163188v1>.
9. Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. 2020. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect Genet Evol* 85:104445. <https://doi.org/10.1016/j.meegid.2020.104445>.
 10. Patro LPP, Sathyaseelan C, Uttamrao PP, Rathinavelan T. 2020. Global variation in the SARS-CoV-2 proteome reveals the mutational hotspots in the drug and vaccine candidates. bioRxiv <https://www.biorxiv.org/content/10.1101/2020.07.31.230987v3>.
 11. Singer J, Gifford R, Cotten M, Robertson D. 2020. CoV-GLUE: a Web application for tracking SARS-CoV-2 genomic variation. Preprints <https://doi.org/10.20944/preprints202006.0225.v1>.
 12. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 1:33–46. <https://doi.org/10.1002/gch2.1018>.
 13. Shu Y, McCauley J. 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22:30494. <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2017.22.13.30494>.
 14. Aiewsakun P, Wongtrakoongate P, Thawornwattana Y, Hongeng S, Thitithanyanont A. 2020. SARS-CoV-2 genetic variations associated with COVID-19 severity. medRxiv <https://www.medrxiv.org/content/10.1101/2020.05.27.20114546v1>.
 15. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. 2020. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet* 65:1075–1082. <https://doi.org/10.1038/s10038-020-0808-9>.
 16. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, Jia X, Wu M, Shi B, Xu S, Chen J, Wang W, Chen B, Jiang L, Yu S, Lu J, Wang J, Xu M, Yuan Z, Zhang Q, Zhang X, Zhao G, Wang S, Chen S, Lu H. 2020. Viral and host factors related to the clinical outcome of COVID-19. *Nature* 583:437–440. <https://doi.org/10.1038/s41586-020-2355-0>.