



SARS-CoV-2 Early Screening at the Point of Entry: Travelers From Bangladesh to Italy–July 2020

Martina Rueca, Antonino Di Caro, Cesare Ernesto Maria Gruber, Francesco Messina, Emanuela Giombini, Maria Beatrice Valli, Eleonora Lalle, Simone Lanini, Francesco Vairo, Maria Rosaria Capobianchi and Barbara Bartolini*

National Institute for Infectious Diseases, INMI (Istituto Nazionale per le Malattie Infettive), “Lazzaro Spallanzani” IRCCS (Istituto di ricovero e Cura a Carattere Scientifico), Rome, Italy

OPEN ACCESS

Edited by:

Sheikh A. Rahman,
Emory University, United States

Reviewed by:

Najmul Haider,
Royal Veterinary College (RVC),
United Kingdom
Mahmuda Yasmin,
University of Dhaka, Bangladesh
Rebecca Rockett,
The University of Sydney, Australia

*Correspondence:

Maria Rosaria Capobianchi
maria.capobianchi@inmi.it

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 03 November 2020

Accepted: 18 January 2021

Published: 09 February 2021

Citation:

Rueca M, Di Caro A, Gruber CEM,
Messina F, Giombini E, Valli MB,
Lalle E, Lanini S, Vairo F,
Capobianchi MR and Bartolini B
(2021) SARS-CoV-2 Early Screening
at the Point of Entry: Travelers From
Bangladesh to Italy–July 2020.
Front. Genet. 12:625607.
doi: 10.3389/fgene.2021.625607

We report phylogenetic and mutational analysis by NGS of six SARS-CoV-2 strains from patients flying from Bangladesh to Italy (July 2020). Data suggest that no further circulation of such imported strains occurred in Italy, stating the efficacy of early screening at the point of entry and supporting the importance of molecular epidemiology in monitoring the efficacy of control measures.

Keywords: SARS-CoV-2, next generation genome sequencing, mutations, phylogenetic analysis, COVID-19, molecular epidemiology, early detection at point of entry

INTRODUCTION

The current outbreak of novel coronavirus (COVID-19) disease has spread across borders through travelers. Thanks to the lockdown measures, closure of unnecessary activities and services, as well as block of traveling from foreign countries was undertaken in Italy from 9th March to 3rd May. As a result, partial control of virus spread was achieved, and subsequently the country gradually returned activities to normal, including travel connections with other countries (ECDC., 2020). In the release phase, early detection of suspect cases at points of entry (POE), including ports, airports and ground crossings and implementation of appropriate control measures are crucial to reduce the risk for igniting new transmission chains (Alm et al., 2020; ECDC., 2020).

Here we report the phylogenetic and mutational analysis of SARS-CoV-2 strains harbored by travelers entering Italy from Bangladesh flying from Dhaka in early July 2020. The analyses were carried out on six samples randomly selected, and the data supported the importance of molecular epidemiology in achieving successful control of new infection waves from imported cases.

METHOD

Nasopharyngeal-swabs (NPS) from 406 travelers coming from Dhaka on 2 airplanes landed in Rome Fiumicino airport were collected immediately upon disembarkment on July 7th, and sent to the Laboratory of Virology of the “L. Spallanzani” Institute, Rome, for SARS-Cov-2 diagnosis, resulting in 50 laboratory confirmed infections.

The presence of SARS-CoV-2 RNA in clinical samples was detected by a commercial RT-PCR assay [Cobas® SARS-CoV-2 (Roche Diagnostics)].

For sequence analysis, the full genome viral sequencing was performed for available residual samples from 6 patients involved in this outbreak. Nucleic acid extraction was performed by QiaSymphony automatic extractor, then Next Generation Sequencing (NGS) was carried out

on Ion Torrent GSS5 platform using Ion AmpliSeq SARS-CoV-2 Panel, following manufacturer's instructions (ThermoFisher). Ethical approval for sequence analysis: no. 70/2018(17/12/2018).

Mean quality Phred score >20 raw reads were selected and trimmed using Trimmomatic software v.0.36 (Bolger et al., 2014).

SARS-CoV-2 genomes were assembled using reference-based assembly method, with BWA v.0.7.12 (Li and Durbin, 2009) and Samtools v.1.3.1 (Li et al., 2009). Contigs were then verified using Geneious 2019.2.3. Single Nucleotide Variants (SNV) were called taking all mutations with a coverage ≥ 50 reads and a frequency

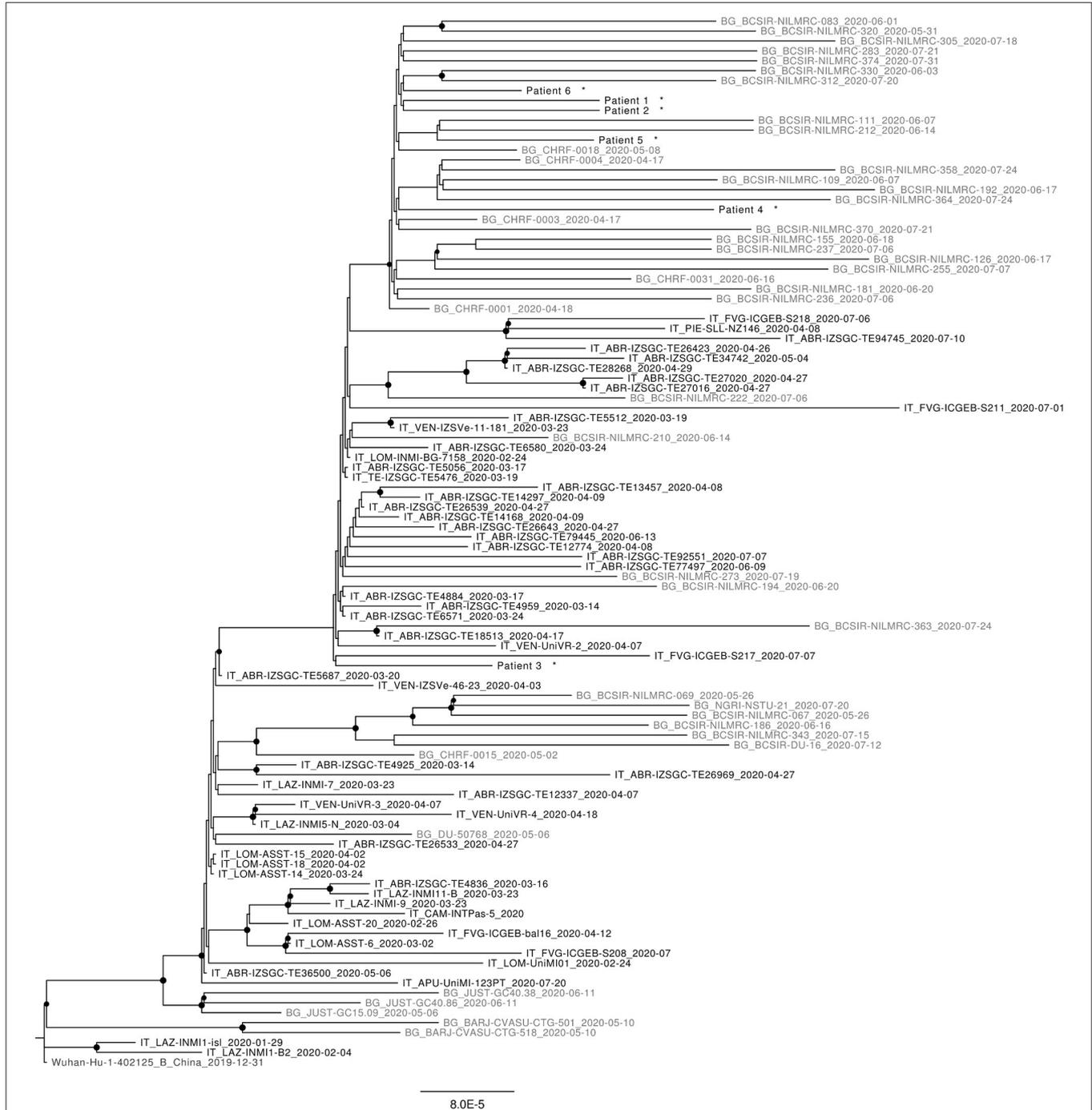


FIGURE 1 | Maximum likelihood phylogenetic tree of representative sequences from Italy (in black) and Bangladesh (in gray) available at 5th October 2020. All Nodes with bootstrap values >65% are highlighted. Sequence from this work are marked with *.

>50%, and excluding mutations lying only in the first or last 5 nucleotides of the reads.

Sequences of SARS-CoV-2 strains from Italy and Bangladesh available at 5th October 2020 were retrieved from GISAID, selecting high coverage genomes. Sequences were clustered at 0.03% using CD-HIT v.4.6 software (Fu et al., 2012). Maximum likelihood phylogenetic analysis was performed with IQ-TREE v.1.6.12, using Transition with invariable sites plus discrete Gamma model (TIM2+I+G) and 1,000 replicates; Wuhan-Hu-1 strain was adopted as phylogenetic outgroup (MN908947.3).

RESULTS

The phylogenetic lineage classification proposed by Rambaut et al. (2020) was used in the phylogenetic analysis, although maintaining, for comparison to previously published reports, also reference to clades reported in GISAID (Elbe and Buckland-Merrett, 2017).

As can be seen in **Figure 1**, most sequences reported from Bangladesh, retrieved from the GISAID platform, form distinct clusters within B1 and B1.1 (G and GR) clades.

TABLE 1 | Consensus sequences of study samples: differences vs. Wuhan-Hu-1 sequence.

Region	nt Position	Ref. *	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	AA change
5'UTR	241	C	T	T	T	T	T	T	–
Orf1ab	829	C				T			Syn
	1,163	A	T	T		T	T	T	I300F
	1,202	A			G				N313D
	3,037	C	T	T	T	T	T	T	Syn
	3,403	G	T						Syn
	4,331	C				T			Syn
	5,281	C			T				Syn
	5,800	G		T					L1845F
	5,869	C						T	Syn
	6,019	A	G						Syn
	7,743	G				T			S2493I
	8,026	A				G			Syn
	8,560	T		C					Syn
	9,136	G		T					M2957I
	10,626	C				T			A3454V
	11,083	G					T		L3606F
	11,719	G			A				Syn
	14,408	C	T	T	T	T	T	T	P4715L
	15,656	C				T			T5131I
	15,807	A						G	Syn
17,427	G					T		Syn	
19,656	G		T					K6464N	
S	23,403	A	G	G	G	G	G	G	D614G
	25,046	C		T					P1162S
	25,047	C					A		P1162C
	25,337	G	T						D1259Y
Orf3a	25,907	G				T		G172V	
Orf7a	27,684	C					T		Syn
	27,707	C						T	A105V
N	28,083	G	T						Syn
	28,881	G	A	A	A	A	A	A	R203K
	28,882	G	A	A	A	A	A	A	R203K
	28,883	G	C	C	C	C	C	C	G204R
	29,272	C			T				Syn
	29,348	G				T			S359S
3'UTR	29,774	C					T		–
	29,779	G	T						–

*Nucleotide positions are referred to Wuhan-Hu-1 (reference genome MN908947).

Nt, nucleotide; AA, amino acid; Syn, synonymous substitution; utr, untranslated region; Orf, open reading frame.

Concerning the study sequences from passengers flying from Bangladesh, all belong to B.1.1 (GR) clade; 5 out of 6 of them fall within the Bangladesh-specific GR cluster highlighted by the phylogenetic analysis; the remaining sequence (sequence number = 3) is interspersed with other GR sequences of mixed origin, among which there are also sequences obtained in Italy in the same period.

Thirty-eight Single Nucleotide Variants (SNV), as compared to the reference strain Wuhan (Accession Number: MN908947), were observed in the study sequences: three in the Untranslated Regions (UTR), 15 synonymous and 20 non-synonymous (Table 1).

More in details, all the strains carry a common set of 7 SNVs: C241T in 5'UTR and C3037T synonymous substitution in ORF1ab, that are the two most abundant mutations found in Bangladesh sequences and often found simultaneously, according to Ahmed Shishir et al. (2020). 14408 C>T and 23403 A>G are two additional non-synonymous mutations, often found simultaneously, leading to P4715L in ORF1ab and D614G changes in Spike protein (the signature mutation for G clade). Finally, the SNVs 28881 G>A, 28882 G>A and 28883 G>C, leading to R203K and G204R changes in Nucleocapsid protein.

Interestingly the non-synonymous SNV 1163 A>T (I300F) in ORF1 is detected in all sequences here described except in patient 3, that is not included in the Bangladesh-specific GR cluster.

DISCUSSION

As emerged from a previous study, a high percentage of virus sequences isolated in India and Bangladesh are closely related to European and US sequences carrying the mutations typical of the G clade (D614G in S and P4715L in Nsp12) (Islam et al., 2020). Moreover, the sequences reported here carry some of the mutations highly prevalent in Bangladesh sequences available on GISAID, among which the I300F mutation (Ahmed Shishir et al., 2020). This is found to affect the structural stability of Nsp2 (methyltransferase like domain), modulating host cell survival strategy (Ahmed Shishir et al., 2020), and deserves further attention.

Since the phylogenetic tree includes all the sequences available by the 5th October from Bangladesh and Italy (0.2% of positive cases in Italy were sequenced at that time), the confinement of most ($n = 5$) sequences within the Bangladesh-specific GR cluster suggests effective containment and despite a limit of this study is represented by the limited number of genomes analyzed (12% of positive cases detected in two flights), no further circulation of virus imported from this country occurred after its importation on 7th of July to Rome.

In a context where travel and business relationships between different countries/continents favor the spread of the virus, sequencing and phylogenetic analysis allows to clearly recognize the cluster of imported sequences, showing strong genetic links with other sequences from the country of origin, and no

further circulation in the country of destination. Therefore, this data show how sequencing of whole-genome of SARS-CoV-2 and phylogenetic analysis are of great support to molecular epidemiology. These results additionally may provide accurate information about the fate of imported viral strains of the virus, hence inform about the efficacy of implemented control measures.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.gisaid.org/>, EPI_ISL_590693; <https://www.gisaid.org/>, EPI_ISL_590694; <https://www.gisaid.org/>, EPI_ISL_590695; <https://www.gisaid.org/>, EPI_ISL_590696; <https://www.gisaid.org/>, EPI_ISL_590697; <https://www.gisaid.org/>, EPI_ISL_590698.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ethics committee of INMI (Ethical approval: no. 70/2018(17/12/2018)). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BB and MR: conceptualization. MR, MV, and EL: methodology. CG and FM: software. CG and EG: formal analysis. MR: investigation. FV and SL: resources. MR: writing—original draft preparation. BB and MC: writing—review and editing. CG: visualization. BB: supervision. MC and AD: project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by funds to the Istituto Nazionale per le Malattie Infettive (INMI) Lazzaro Spallanzani IRCCS, Rome, Italy, from the Ministero della Salute (Ricerca Corrente, linea 1; COVID-2020-12371817), the European Commission—Horizon 2020 (EU project 101003544—CoNVat; EU project 101005075—KRONO) and the European Virus Archive—GLOBAL (Grants Nos. 653316 and 871029).

ACKNOWLEDGMENTS

We gratefully acknowledge the contributors of genome sequences of the newly emerging coronavirus, i.e., the Originating and Submitting Laboratories, for sharing their sequences and other metadata through the GISAID Initiative.

REFERENCES

- Ahmed Shishir, T., Bin Naser, I., and Faruque, S. M. (2020). In silico comparative genomics of SARS-CoV-2 to determine the source and diversity of the pathogen in Bangladesh. *bioRxiv*. doi: 10.1101/2020.07.20.212563
- Alm, E., Broberg, E. K., Connor, T., Hodcroft, E. B., Komissarov, A. B., Maurer-Stroh, S., et al. (2020). Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European region, January to June 2020. *Euro Surveill.* 25:2001410. doi: 10.2807/1560-7917.ES.2020.25.32.2001410
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- ECDC. (2020). *Coronavirus Disease 2019 (COVID-19) in the EU/EEA and the UK – Tenth Update What is New in This Update? What Are the Risks Being Assessed in this Update?* European Centre for Disease Prevention and Control.
- Elbe, S., and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Challen.* 1, 33–46. doi: 10.1002/gch2.1018
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Islam, O. K., Al-Emran, H. M., Hasan, M. S., Anwar, A., Jahid, M. I. K., and Hossain, M. A. (2020). Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia. *Transbound Emerg Dis.* 1–9. doi: 10.1111/tbed.13748
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Rambaut, A., Holmes, E. C., Hill, V., O'Toole, A., McCrone, J., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv*. doi: 10.1101/2020.04.17.046086

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Rueca, Di Caro, Gruber, Messina, Giombini, Valli, Lalle, Lanini, Vairo, Capobianchi and Bartolini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.