Isolation and full-length genome characterization of SARS-CoV-2 from COVID-19 cases in Northern Italy

[1,#]Danilo Licastro, [2,#]Sreejith Rajasekharan, [1]Simeone Dal Monego, [3]Ludovica Segat, [3]Pierlanfranco D'Agaro, [2]Alessandro Marcello and [4]The Regione FVG Laboratory Group on COVID-19.

[1]ARGO Open Lab Platform for Genome sequencing, AREA Science Park, Padriciano, 99, 34149, Trieste, Italy.

[2]Laboratory of Molecular Virology, International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano, 99, 34149, Trieste, Italy.

[3]Laboratorio di riferimento per SARS-CoV-2, Regione Friuli-Venezia Giulia, Azienda Sanitaria Universitaria Integrata Giuliano Isontina (ASUGI), UCO Igiene e Sanità Pubblica, Dipartimento di Scienze Mediche Chirurgiche e della Salute, Università di Trieste, Via della Pietà,2/2, 34129 Trieste, Italy.

[4]The members of the Regione FVG laboratory Group on COVID-19 are listed at the end of the article.

[#]These authors contributed equally to this work and are named in alphabetical order.

*To whom correspondence may be addressed: AM marcello@icgeb.org, PDA pdagaro@units.it

In December 2019, the novel coronavirus Severe Acquired Respiratory Syndrome SARS-CoV-2 emerged in the city of Wuhan in the Hubei province, People's Republic of China, as the etiologic agent of coronavirus disease 2019 (COVID-19), which has hence spread worldwide causing a global pandemic (1-3). The epidemic has been growing exponentially in Italy for the last month affecting over 60.000 individuals so far and with a heavy mortality burden. Italy is only anticipating what will be the trend in whole Europe and elsewhere. At the beginning of March 2020, the first nasopharyngeal swabs positive for SARS-CoV-2 started to be detected in the Northern Eastern Region of Friuli Venezia Giulia. These identifications followed the expansion of the two clusters in Lombardy and Veneto that emerged the previous weeks in Northern Italy (4). Swabs contents were seeded on Vero E6 cells, and monitored for cytopathic effect and by an RT PCR protocol using primers for the N region (5). Cell culture supernatant from passage 1 (P1) of four isolates were collcted, and RNA was extracted with QiAmp Viral RNA mini kit (Qiagen), and quantified

with an *in vitro* transcribed RNA standard (6). The quantity and quality of the RNA was assessed using Qubit 2.0 fluorometer (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies). For each sample 100 ng of total RNA was processed using Zymo-Seq RiboFree Ribosomal depletion library preparation kit (Zymo Research). All the obtained libraries passed quality check and were quantified before being pooled at equimolar concentration and sequenced on Illumina Nano MiSEQ 2x150bp paired-end mode following standard procedures. Sequenced reads that passed the quality check (Phred score ≥30), were adaptor and quality trimmed, the remaining reads were assembled de novo using Megahit (v.1.2.9) with default parameter settings. Megahit generated in all cases 7 contigs with more than 1000bp and 100x coverage, all of these assembled contigs were compared (using BLASTn) against the entire non-redundant (nr) nucleotide and protein databases. In all cases the longest and more covered contigs was identifies as MT019532.1 “Severe acute respiratory syndrome coronavirus 2 isolate BetaCoV/Wuhan/IPBCAMS-WH-04/2019, complete genome“ with 99% of identity and 0 gaps. The longer sequences were named hCoV-19/Italy/FVG/ICGEB_S1, _S5, S8, S9 and were deposited in GISAID with accession numbers EPI_ISL_417418, EPI_ISL_417419, EPI_ISL_417421 and EPI_ISL_417423, respectively (7). Sequence analysis showed an uneven coverage along the SARS-CoV-2 genome, with an average range from 126 to 7576 reads and a mean coverage per sample of 1169x (Figure 1). Phylogenetic trees were inferred using the maximum likelihood method implemented in the MEGAX program using the GISAID sequences available at 03-16-2020 (8). Bootstrap support values were calculated from 500 pseudo-replicate trees of the whole dataset (Figure 2).

Despite a high burden of COVID-19 in Italy, very little information is available to date from full-length high quality sequences. The first sequences deposited on GISAID (EPI_ISL_410545 and EPI_ISL_410546) were collected in Rome from a Chinese tourist from the Hubei province who got infected before visiting Italy and another one (EPI_ISL_412974) from a positive Italian citizen returning from China. Only two sequence were reported from the Lombardy cluster (EPI_ISL_412973 and EPI_ISL_413489). In this report four additional sequences from cases epidemiologically linked to Northern Italy have been examined. All infected individuals were related to the city of Udine, S1 and S5 were from the same cluster of closely related cases, while S9 got infected probably in Lombardy and S8 visited Udine from a neighbouring city (Table 1). Sequence analysis showed a good coverage along the SARS-CoV-2 genome for all four isolates (Figure 1). Based on the marker variant S D614G, all four sequences grouped in the Bavarian rooted subclade G, which is dominant in Europe, including the sequence from Lombardy, but distinct form the three sequences mentioned above originating directly from China (9). Intriguingly, the

new isolates were more closely related to EPI_ISL_412973, while EPI_ISL_413489 was more distant (Figure 2). No evidence could be found for the putative 382-nt deletion in ORF8 detected in Singapore, which has been proposed to indicate an attenuated phenotype (10).

These findings strongly urge the need for comprehensive studies that combine genomic data with epidemiological data and clinical records of symptoms from patients with COVID-19.

[4]*The Regione FVG laboratory Group on COVID-19.*

Gianna Zamaro, Direzione centrale salute della Regione Friuli-Venezia Giulia, Trieste, Italy

Tolinda Gallo, Direzione centrale salute della Regione Friuli-Venezia Giulia, Trieste, Italy

Raffaella Koncan, Laboratorio di riferimento per SARS-CoV-2, Regione Friuli-Venezia Giulia, Trieste, Italy

Maurizio Ruscio, Dipartimento dei Servizi, ASUGI, Trieste, Italy

Corrado Pipan, Laboratorio di virologia, ASUFC, Udine, Italy

Francesco Curcio, Dipartimento di Laboratorio, ASUFC, Udine, Italy

Rita de Rosa, SC Microbiologia e Virologia, ASFO, Pordenone, Italy

Acknowledgments

Work financed through intramural funds of the ICGEB, University of Trieste, ASUGI and the ARGO, MIUR-MISE-Region FVG Agreement Protocol signed by MIUR-MISE-Region FVG and managed by Area Science Park. We thank Elena Samar, Fabia Petronio and Ilaria Sandron for their excellent technical assistance.

References

1. **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.
2. **Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ.** 2020. A new coronavirus associated with human respiratory disease in China. Nature **579:**265-269.
3. **WHO.** 2020. World Health Organization - Coronavirus disease 2019 (COVID-19). Situation Report – 61. https://wwwwhoint/emergencies/diseases/novel-coronavirus-2019/situation-reports/.
4. **Anonymous.** 2020. Dipartimento della Protezione Civile - Emergenza Coronavirus: la risposta nazionale. http://www.protezionecivile.gov.it/attivitarischi/rischiosanitario/emergenze/coronavirus.

5. **Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, Bleicker T, Brunink S, Schneider J, Schmidt ML, Mulders D, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MPG, Drosten C.** 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill **25**.
6. **Rajasekharan S, Marcello A.** 2020. Multitarget standard RNA for SARS-CoV-2 diagnosis. wwwicgeborg **(manuscript in preparation)**.
7. **Shu Y, McCauley J.** 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill **22**.
8. **Kumar S, Stecher G, Li M, Knyaz C, Tamura K.** 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol **35:**1547-1549.
9. **Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M.** 2020. The first two cases of 2019-nCoV in Italy: Where they come from? J Med Virol **92:**518-521.
10. **Su YCF, Anderson DE, Young BE, Zhu F, Linster M, Kalimuddin S, Low JGH, Yan Z, Jayakumar J, Sun L, Yan GZ, Mendenhall IH, Leo Y-S, Lye DC, Wang L-F, Smith GJD.** 2020. Discovery of a 382-nt deletion during th 1 e early evolution of SARS-CoV-2. https://wwwbiorxivorg/content/101101/20200311987222v1.

Figure legends

Figure 1

Samples Coverage tracks from UCSC Browser on ASM985889v3/SARS-CoV-2 Assembly. Loaded tracks include UniProt Proteins track, RefSeq Acc track and samples coverage tracks obtained after mapping raw reads to ASM985889v3 and converted using bedtools genomecov function. Highlighted position refers to D614G variation in S protein revealed in all sequenced cases.

Figure 2

Maximum likelihood phylogenetic trees of nucleotide sequences from GISAID sequences available at 03-16-2020 and hCoV-19/Italy/FVG/ICGEB_S1, S5, S8, _S9. A portion of the G clade based on S variation D614G is shown with indication of the phylogenetic tree branches including reporting cases (purple) and the other two deposited Lombardy sequences (red dot).

Table 1

SEQUENCE	AGE	SEX	RESIDENCE	CLUSTER
#1	59	F	Udine	Meeting in Udine
#5	73	M	Udine	Meeting in Udine
#8	45	M	Gorizia	Visit to Udine
#9	59	M	Udine	Visit to Milan

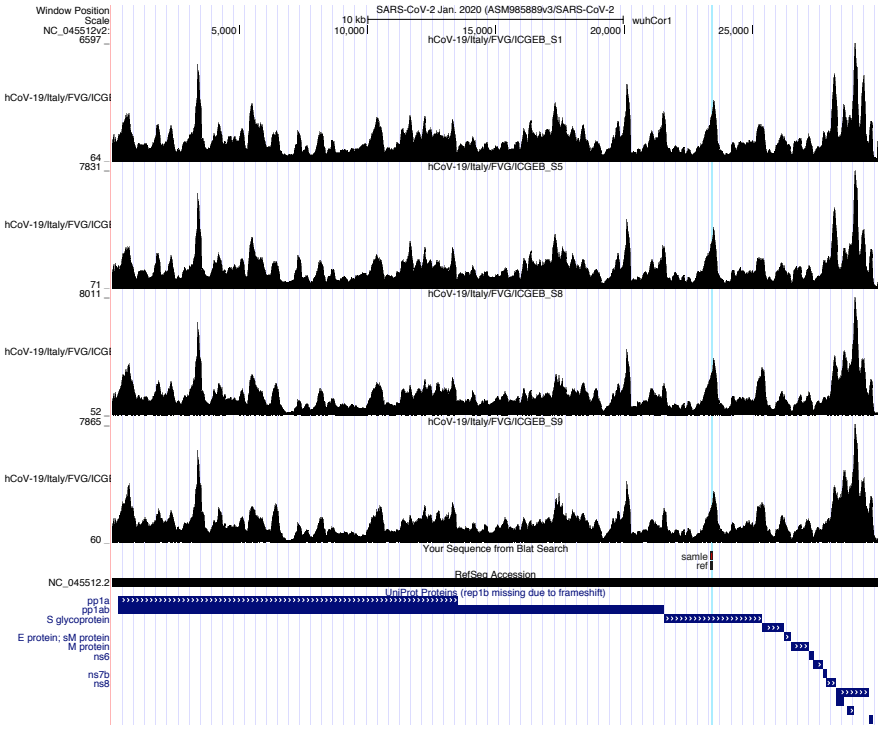

Figure 1

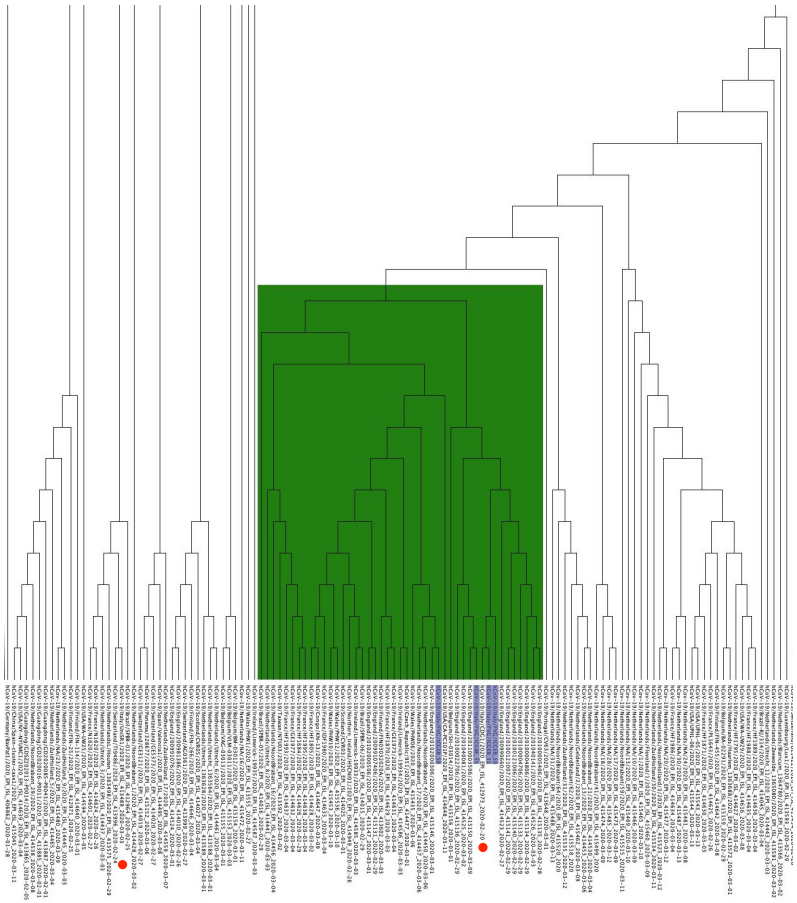

Figure 2