

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

3
4 ^{1,‡}Danilo Licastro, ^{2,‡}Sreejith Rajasekharan, ¹Simeone Dal Monego, ³Ludovica Segat,
5 ³Pierlanfranco D'Agaro, ²Alessandro Marcello and ⁴The Regione FVG Laboratory Group
6 on COVID-19.

7
8 ¹ARGO Open Lab Platform for Genome sequencing, AREA Science Park, Padriciano, 99,
9 34149, Trieste, Italy.

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

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

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

18
19 [‡]These authors contributed equally to this work and are named in alphabetical order.

20
21 *To whom correspondence may be addressed: AM marcello@icgeb.org, PDA
22 pdagaro@units.it

23
24 In December 2019, the novel coronavirus Severe Acquired Respiratory Syndrome SARS-
25 CoV-2 emerged in the city of Wuhan in the Hubei province, People's Republic of China, as
26 the etiologic agent of coronavirus disease 2019 (COVID-19), which has hence spread
27 worldwide causing a global pandemic (1-3). The epidemic has been growing exponentially
28 in Italy for the last month affecting over 60.000 individuals so far and with a heavy mortality
29 burden. Italy is only anticipating what will be the trend in whole Europe and elsewhere. At
30 the beginning of March 2020, the first nasopharyngeal swabs positive for SARS-CoV-2
31 started to be detected in the Northern Eastern Region of Friuli Venezia Giulia. These
32 identifications followed the expansion of the two clusters in Lombardy and Veneto that
33 emerged the previous weeks in Northern Italy (4). Swabs contents were seeded on Vero
34 E6 cells, and monitored for cytopathic effect and by an RT PCR protocol using primers for
35 the N region (5). Cell culture supernatant from passage 1 (P1) of four isolates were
36 collected, and RNA was extracted with QiAmp Viral RNA mini kit (Qiagen), and quantified

37 with an *in vitro* transcribed RNA standard (6). The quantity and quality of the RNA was
38 assessed using Qubit 2.0 fluorometer (Thermo Fisher Scientific) and Agilent 2100
39 Bioanalyzer (Agilent Technologies). For each sample 100 ng of total RNA was processed
40 using Zymo-Seq RiboFree Ribosomal depletion library preparation kit (Zymo Research).
41 All the obtained libraries passed quality check and were quantified before being pooled at
42 equimolar concentration and sequenced on Illumina Nano MiSEQ 2x150bp paired-end
43 mode following standard procedures. Sequenced reads that passed the quality check
44 (Phred score ≥ 30), were adaptor and quality trimmed, the remaining reads were
45 assembled de novo using Megahit (v.1.2.9) with default parameter settings. Megahit
46 generated in all cases 7 contigs with more than 1000bp and 100x coverage, all of these
47 assembled contigs were compared (using BLASTn) against the entire non-redundant (nr)
48 nucleotide and protein databases. In all cases the longest and more covered contigs was
49 identifies as MT019532.1 “Severe acute respiratory syndrome coronavirus 2 isolate
50 BetaCoV/Wuhan/IPBCAMS-WH-04/2019, complete genome“ with 99% of identity and 0
51 gaps. The longer sequences were named hCoV-19/Italy/FVG/ICGEB_S1, _S5, S8, S9 and
52 were deposited in GISAID with accession numbers EPI_ISL_417418, EPI_ISL_417419,
53 EPI_ISL_417421 and EPI_ISL_417423, respectively (7). Sequence analysis showed an
54 uneven coverage along the SARS-CoV-2 genome, with an average range from 126 to
55 7576 reads and a mean coverage per sample of 1169x (Figure 1). Phylogenetic trees
56 were inferred using the maximum likelihood method implemented in the MEGAX program
57 using the GISAID sequences available at 03-16-2020 (8). Bootstrap support values were
58 calculated from 500 pseudo-replicate trees of the whole dataset (Figure 2).
59 Despite a high burden of COVID-19 in Italy, very little information is available to date from
60 full-length high quality sequences. The first sequences deposited on GISAID
61 (EPI_ISL_410545 and EPI_ISL_410546) were collected in Rome from a Chinese tourist
62 from the Hubei province who got infected before visiting Italy and another one
63 (EPI_ISL_412974) from a positive Italian citizen returning from China. Only two sequence
64 were reported from the Lombardy cluster (EPI_ISL_412973 and EPI_ISL_413489). In this
65 report four additional sequences from cases epidemiologically linked to Northern Italy have
66 been examined. All infected individuals were related to the city of Udine, S1 and S5 were
67 from the same cluster of closely related cases, while S9 got infected probably in Lombardy
68 and S8 visited Udine from a neighbouring city (Table 1). Sequence analysis showed a
69 good coverage along the SARS-CoV-2 genome for all four isolates (Figure 1). Based on
70 the marker variant S D614G, all four sequences grouped in the Bavarian rooted subclade
71 G, which is dominant in Europe, including the sequence from Lombardy, but distinct from
72 the three sequences mentioned above originating directly from China (9). Intriguingly, the

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

77 These findings strongly urge the need for comprehensive studies that combine genomic
78 data with epidemiological data and clinical records of symptoms from patients with COVID-
79 19.

80

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

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

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

84 *Raffaella Koncan, Laboratorio di riferimento per SARS-CoV-2, Regione Friuli-Venezia*

85 *Giulia, Trieste, Italy*

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

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

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

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

90

91 **Acknowledgments**

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

96

97 **References**

- 98 1. **Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang**
99 **CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang**
100 **X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao**
101 **GF, Shi ZL.** 2020. A pneumonia outbreak associated with a new coronavirus of
102 probable bat origin. *Nature* **579**:270-273.
- 103 2. **Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei**
104 **YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC,**
105 **Zhang YZ.** 2020. A new coronavirus associated with human respiratory disease in
106 China. *Nature* **579**:265-269.
- 107 3. **WHO.** 2020. World Health Organization - Coronavirus disease 2019 (COVID-19).
108 Situation Report – 61. [https://www.who.int/emergencies/diseases/novel-coronavirus-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/)
109 [2019/situation-reports/](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/).
- 110 4. **Anonymous.** 2020. Dipartimento della Protezione Civile - Emergenza
111 Coronavirus: la risposta nazionale.
112 <http://www.protezionecivile.gov.it/attivitarischi/rischiosanitario/emergenze/coronavirus>.

- 113 5. **Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, Bleicker T,**
 114 **Brunink S, Schneider J, Schmidt ML, Mulders D, Haagmans BL, van der Veer B,**
 115 **van den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris**
 116 **M, Goossens H, Reusken C, Koopmans MPG, Drosten C.** 2020. Detection of 2019
 117 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* **25**.
 118 6. **Rajasekharan S, Marcello A.** 2020. Multitarget standard RNA for SARS-CoV-2
 119 diagnosis. *www.icgeb.org (manuscript in preparation)*.
 120 7. **Shu Y, McCauley J.** 2017. GISAID: Global initiative on sharing all influenza data -
 121 from vision to reality. *Euro Surveill* **22**.
 122 8. **Kumar S, Stecher G, Li M, Knyaz C, Tamura K.** 2018. MEGA X: Molecular
 123 Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**:1547-
 124 1549.
 125 9. **Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M.** 2020. The first two cases of
 126 2019-nCoV in Italy: Where they come from? *J Med Virol* **92**:518-521.
 127 10. **Su YCF, Anderson DE, Young BE, Zhu F, Linster M, Kalimuddin S, Low JGH,**
 128 **Yan Z, Jayakumar J, Sun L, Yan GZ, Mendenhall IH, Leo Y-S, Lye DC, Wang L-F,**
 129 **Smith GJD.** 2020. Discovery of a 382-nt deletion during the early evolution of
 130 SARS-CoV-2. <https://www.biorxiv.org/content/101101/20200311987222v1>.
 131

132 Figure legends

133 Figure 1

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

140 Figure 2

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

147 Table 1

148

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

149

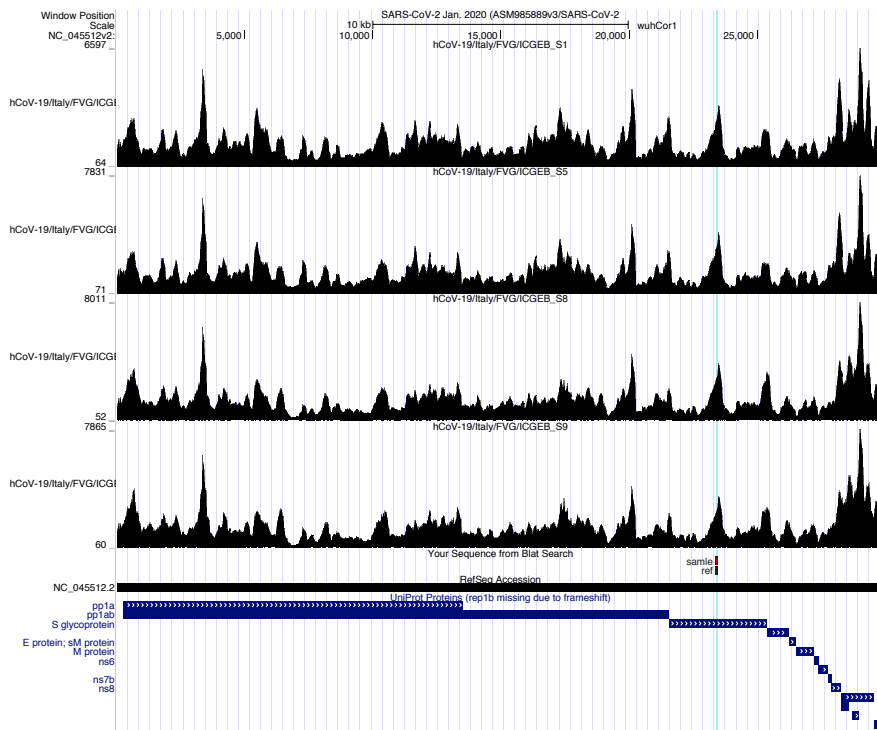


Figure 1

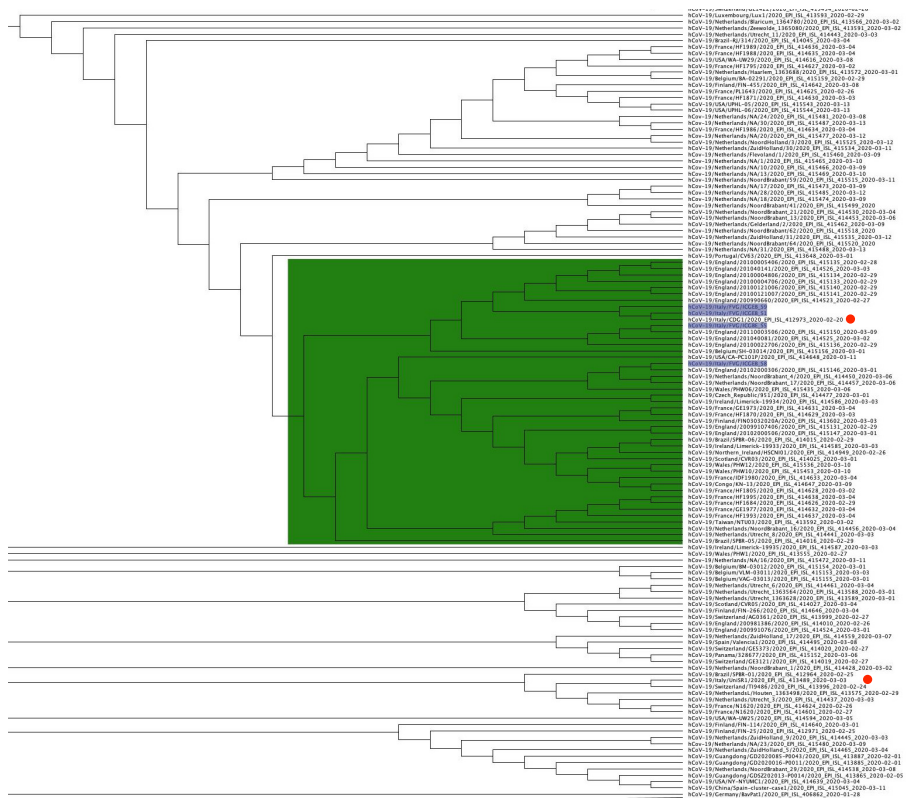


Figure 2