Clinical Microbiology and Infection 26 (2020) 954–956

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Letter to the Editor

Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy

M.R. Capobianchi, M. Rueca, F. Messina, E. Giombini, F. Carletti, F. Colavita, C. Castilletti, E. Lalle, L. Bordi, F. Vairo, E. Nicastri, G. Ippolito, C.E.M. Gruber*, B. Bartolini

National Institute for Infectious Diseases Lazzaro Spallanzani IRCCS, Rome, Italy

ARTICLE INFO

Article history:
Received 26 February 2020
Received in revised form
17 March 2020
Accepted 21 March 2020
Available online 27 March 2020

Editor: L. Leibovici

To the Editor,

On January 29, 2020, two Chinese spouses (patient 1, female; patient 2, male), coming to Italy as tourists from Hubei province, were hospitalized at the National Institute for Infectious Diseases "L. Spallanzani", Rome, with fever and respiratory symptoms. SARS-CoV-2 diagnosis was accomplished using real-time RT-PCR [1] on a nasopharyngeal swab and sputum for patient 1 and on a nasopharyngeal swab for patient 2, collected 1 day after symptom onset. Partial sequencing confirmed both patients to be infected with SARS-CoV-2.

A virus isolate was obtained (in a Vero E6 cell line) from the sputum of patient 1, with cytopathic effects evident 24 h post-inoculation. At the time of writing, virus isolation from the nasopharyngeal swab sample collected from patient 2 was not successful, likely due to the lower viral load (higher cycle threshold value, 24.56 in the real-time RT-PCR), therefore no further analysis was performed on the virus detected in patient 2. Next-generation sequencing (NGS) was performed on the respiratory samples from patient 1 and on the primary isolate, prior to any further passage, by using the Ion Torrent S5 platform (Thermofisher). The mean count of sequencing reads obtained per sample was 44 000 000 (minimum 41.6 × 10^6 to maximum 49.7 × 10^6). The reads from the two respiratory samples of patient 1 were merged to obtain a better coverage along the virus genome, and in this paper are referred to as data from the clinical sample. Details of sequencing and bioinformatic analyses are available upon request.

The number of SARS-CoV-2 reads obtained varied from 4079 to >14 × 10^6. By using *de novo* assembly, two contigs of 29 867 nt (mean coverage: 81 324 reads; range: 26–510 718 reads) and 29 792 nt (mean coverage: 80 reads; range: 5–599 reads) were obtained for the isolate and clinical sample of patient 1, respectively, and referred to as consensus sequences. Further analysis was dedicated to identifying the variants present at any nucleotide position for the variability analysis.

Considering the consensus sequences, two non-synonymous changes with respect to the Wuhan-Hu-1 NCBI Reference Genome (Accession number: MN908947.3) [2] were observed in the sequence from the clinical sample from patient 1: G11083T, leading to L3606F change in Orf1a, and G26144T, leading to G251V change in Orf3a. One additional synonymous substitution in Orf1a (A2269T) was detected in the isolate but not in the corresponding clinical sample. All variants were confirmed by Sanger sequencing.

Considering the analysis of genomic variability, several intra-sample variants were observed in both the isolate and the clinical sample, but only the positions with a minimum coverage of 20 reads were considered. Intra-sample assessment of overall virus genome variability resulted in 1.27 × 10^{-4} and 1.02 × 10^{-4} nucleotide substitutions per site for the isolate and the clinical sample, respectively. Only two variable positions were observed with a frequency >10% in the clinical sample, both in Orf1a: A2269T (13.73%, coverage: 51x), synonymous for amino acid A668, and G7388A (13.21%, coverage: 53x), leading to amino acid change (A2375T). Interestingly, the frequency of variants at position 2269 was different in the isolate, being T dominant over A in 72% of reads (coverage: 119 582x), accounting for the difference resulting in the consensus sequences.

For the phylogenetic analysis, 87 full-genome SARS-CoV-2 sequences were retrieved from the Global Initiative on Sharing All Influenza Data (GISAID), along with WH-01_MN908947.3 from GenBank. The G26144T substitution observed in the isolate from Italy was also present in five sequences from cases occurring

* **Corresponding author**. C.E.M. Gruber, Laboratory of Virology, INMI Lazzaro Spallanzani IRCCS, via Portuense 292, 00149, Roma, Italy.
E-mail address: cesare.gruber@inmi.it (C.E.M. Gruber).

https://doi.org/10.1016/j.cmi.2020.03.025

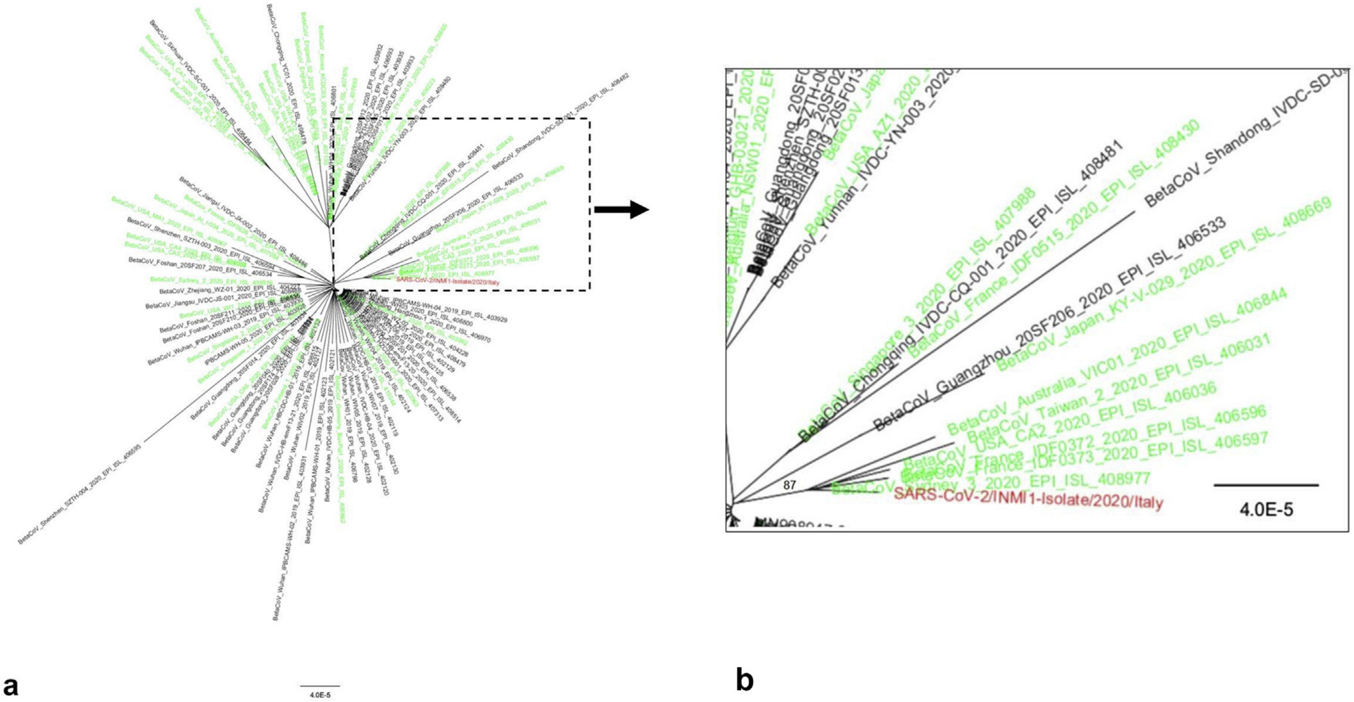
a
4.0E-5
b
BetaCoV_Shandong_IVDC-SD-0
BetaCoV_Chongqing_IVDC-CQ-001_2020_EPI_ISL_408481
BetaCoV_France_IDF0515_2020_EPI_ISL_408430
BetaCoV_Singapore_3_2020_EPI_ISL_407988
BetaCoV_Yunnan_IVDC-YN-003_2020
BetaCoV_USA_AZ1_2020
BetaCoV_Guangzhou_20SF206_2020_EPI_ISL_406533
BetaCoV_Japan_KY-V-029_2020_EPI_ISL_408669
BetaCoV_Australia_VIC01_2020_EPI_ISL_406844
BetaCoV_Taiwan_2_2020_EPI_ISL_406031
BetaCoV_USA_CA2_2020_EPI_ISL_406036
BetaCoV_France_IDF0372_2020_EPI_ISL_406596
BetaCoV_France_IDF0373_2020_EPI_ISL_406597
BetaCoV_Sydney_3_2020_EPI_ISL_408977
SARS-CoV-2/INMI1-Isolate/2020/Italy
87
4.0E-5

Fig. 1. (a) Unrooted phylogenetic tree based on SARS-CoV-2 full genome database from the Global Initiative on Sharing All Influenza Data (GISAID). Genomes collected outside China are highlighted in green. (b) Enlargement of clade reporting SARS-CoV-2/INMI1-Isolate/2020/Italy (in red). Maximum likelihood phylogeny was reconstructed under Hasegawa-Kishino-Yano plus proportion of invariable sites (HKY + I), inferred by model test function.

outside of China: EPI_ISL_406596 and EPI_ISL_406597 from France, EPI_ISL_406031 from Taiwan, EPI_ISL_406036 from USA, EPI_ISL_406844 and EPI_ISL_408977 from Australia. All the genomes carrying this mutation are included in a significant phylogenetic cluster (bootstrap 87%), suggesting a common origin (Fig. 1); in fact, the G251V substitution in Orf3a has recently been defined as the marker variant of the 'V' clade (GISAID).

The presence of quasispecies has previously been reported for SARS-CoV and MERS-CoV [3,4], suggesting that these beta-coronaviruses may consist of complex and dynamic distributions of closely related variants *in vivo*, similarly to other RNA viruses. When applied to SARS-CoV-2 in this study, the analysis of sequence variability supported the presence of viral quasispecies in the clinical sample as well as in the primary isolate. Namely, two positions with variant frequency >10% were observed in the biological sample, both in Orf1a: A2269T, synonymous, and G7388A, corresponding to amino acid change A2375T. The synonymous variant A2269T, representing a minority variant in the clinical sample, was the dominant one in the isolate. Although low coverage may have affected the precise calculation of minority variant frequency in the clinical sample, the data are consistent with variant selection occurring during the isolation procedure, as previously shown for other respiratory viruses. In the respiratory sample neither mutations nor intra-sample variants were found at positions 8782 and 28 144, recently identified as hotspots of hypervariability (coverage: 61x and 76x respectively) [5].

Full-genome characterization of new viruses is instrumental for updating diagnostics and assessing viral evolution. On the other hand, virus variability, leading to the development of quasispecies within infected patients, may provide the background for virus evolution and adaptation to new hosts; more studies are necessary to unravel the importance of intra-patient variability in the SARS-CoV-2 evolutionary trajectory.

Genome sequences described on this manuscript are available from GISAID and from GenBank (Acc. Numb: MT008022, MT008023, MT066156 and MT077125).

Author contributions

All authors contributed to the analysis and the writing of the final manuscript.

Transparency declaration

The authors declare that they have no conflicts of interest. This research was supported by funds to the National Institute for Infectious Diseases 'Lazzaro Spallanzani' IRCCS from the Ministero della Salute, Ricerca Corrente.

Acknowledgements

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, on which this research is based.

References

[1] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 2020;25. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045. pii=2000045.

[2] Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265–9. https://www.nature.com/articles/s41586-020-2008-3.

[3] Park D, Huh HJ, Kim YJ, Son DS, Jeon HJ, Im EH, et al. Analysis of intrapatient heterogeneity uncovers the microevolution of Middle East respiratory syndrome coronavirus. Cold Spring Harbor Mol Case Stud 2016;2:a001214. https://doi.org/10.1101/mcs.a001214.

[4] Xu D, Zhang Z, Wang FS. SARS-associated coronavirus quasispecies in individual patients. New Engl J Med 2004;350:1366–7. https://doi.org/10.1056/NEJMc032421.

[5] Ceraolo C, Giorgi F. Genomic variance of the 2019-nCoV coronavirus. J Med Virol 2020;92:522–8. https://doi.org/10.1002/jmv.25700.