



Silvia Angeletti ORCID iD: 0000-0002-7393-8732

Domenico Benvenuto ORCID iD: 0000-0003-3833-2927

Massimo Ciccozzi ORCID iD: 0000-0003-3866-9239

A doubt of multiple introduction of SARS-CoV-2 in Italy: a preliminary overview

Marta Giovanetti^{1,2}, Silvia Angeletti³, Domenico Benvenuto⁴, Massimo Ciccozzi⁴.

¹Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; ²Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; ³ Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Italy; ⁴ Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Italy.

Corresponding author:

Prof. Silvia Angeletti

Unit of Clinical Laboratory Science

University Campus Bio-Medico of Rome, Italy.

++3906225411461

e-mail: s.angeletti@unicampus.it

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jmv.25773.

This article is protected by copyright. All rights reserved.

Running head: SARS-CoV-2 Italian epidemic

ABSTRACT

The emergence of the novel beta *Coronavirus*, recently renamed as severe acute respiratory syndrome coronavirus 2, SARS-CoV-2, has raised serious concerns due to the virus rapid dissemination worldwide. Nevertheless, there is limited information about the genomic epidemiology of SARS-CoV-2 circulating in Italy from surveillance studies. The shortage of complete genomic sequences available impairs our understanding of the SARS-CoV-2 introduction and establishment in the country. To better understand its dynamics in Italy, we analysed complete genomes of SARS-CoV-2 isolates, obtained directly from clinical samples. Our phylogenetic reconstructions suggest possible multiple introduction of SARS-CoV-2. Continued genomic surveillance strategies are needed to improve monitoring and understanding of the currently SARS-CoV-2 epidemics, which might help to attenuate public health impact of infectious diseases.

Keywords: Italian outbreak, COVID-2019, phylogenetic inference

INTRODUCTION

Coronaviruses (CoVs) are enveloped, positive-sense, single-stranded RNA viruses belonging to the Coronaviridae family¹. They have been categorized based on antigenic properties and phylogenetic analyses in three main groups²: 1) alpha-CoVs, mainly responsible for gastrointestinal disorders; 2) beta-CoVs, that includes: (i) Bat coronavirus (BCoV), (ii) the human Severe Acute Respiratory Syndrome (SARS) virus, (iii) the Middle Eastern Respiratory Syndrome (MERS) virus; 3) gamma-CoVs, that mainly infect avian species. Some variants of coronaviruses are associated with

This article is protected by copyright. All rights reserved.

outbreaks, some others are continuously circulating and cause mostly mild respiratory infections (e.g. the common cold). The most well known of these coronaviruses is the SARS-CoV ("severe acute respiratory syndrome"), which between 2002-2003 was responsible to cause an outbreak that spread around the world and resulted in over 8000 cases and 774 deaths, with a case fatality rate of around 9–11%³. In 2012, a novel coronavirus, MERS-CoV ("Middle East respiratory syndrome"), causing severe respiratory symptoms was identified⁴. Very recently, in December 2019, a novel beta-CoVs Coronavirus, was responsible of a new illness, first detected in Wuhan, China⁵. We now know this to be another outbreak of coronavirus in humans (the 7th), that was recently renamed as severe acute respiratory syndrome coronavirus 2, SARS-CoV-2. The origin of the virus is still unclear, however genomic analysis suggests nCoV is most closely related to viruses previously identified in bats⁶.

As the first week of March 2020 over 98.192 confirmed cases and 3.380 deaths have been reported globally⁷. The case counts are dramatically rising in part due to increased surveillance and testing. While the outbreak seems to be centered in Wuhan, which is now under quarantine, the virus has spread throughout China and abroad, including Western and South-Eastern pacific regions, European regions, Eastern Mediterranean regions as well as many states in the Americas⁸. It comes amid signs of the virus's shift away from its origins in China, with Italy, Iran and South Korea emerging as the other countries hardest hit by the deadly disease. Italy is the worst affected country in Europe, with more than 4,636 cases, a rise of 778 in a day, and a death toll of 197, an increase of 49⁹. Nevertheless, there is still limited information about the genomic epidemiology of the SARS-CoV-2 circulating in Italy from genomic surveillance studies, impairs our understanding of the virus introduction, establishment and dissemination in the country. Thus, to better understand the emergence of the recent SARS-CoV-2 epidemic in Italy, This article is protected by copyright. All rights reserved.

we analyzed a larger and updated dataset of recently released data of the COVID-2019 epidemic in order to improve more effective intervention strategies.

MATERIALS AND METHODS

Phylogenetic analysis

The dataset comprised (n= 141) complete genomes sequences from the current (2019-2020) SARS-CoV-2 epidemic retrieved from GISAID (<https://www.gisaid.org/>) database. Alignment was performed using MAFFT online program¹⁰. The complete dataset was assessed for presence of phylogenetic signal by applying the likelihood mapping analysis implemented in the IQ-TREE 1.6.8 software (<http://www.iqtree.org>)¹¹. A maximum likelihood (ML) phylogeny was reconstructed using IQ-TREE 1.6.8 software under the general time-reversible nucleotide substitution model with a proportion of invariant sites (GTR + I) which was inferred in jModelTest (<https://github.com/ddarriba/jmodeltest2>) as the best fitting model¹².

RESULTS

To better understand the establishment and transmission of the novel SARS-CoV-2 in Italy, the recent Italian genomes available on GISAID database were added to a dataset of 138 SARS-CoV-2 genomes from the global epidemic. A maximum likelihood (ML) phylogenetic tree (**Figure 1**) was built to explore the recent Italian genomes relationship with previous worldwide isolates. Maximum likelihood phylogenies suggested that the three Italian strains, grouped into two different well-supported phylogenetic clades (bootstrap score 0.80; 0.99 respectively). We found that 2 isolate from Italy (EPI_ISL_410546 a Chinese tourist from Hubei Province, and EPI_ISL_412974 the first Italian case isolate in Rome Italy) fall within a clade (bootstrap score, BS = 94%)

(**Figure 1A clade 1**) including sequences from several countries as China, Japan, Singapore, England, Australia and Canada. However, it is important to note that sequences from China are positioned basally to this first clade, reinforcing the idea that the first introduction event could be mediated, as suspected by China.

We also identified that one Italian strain (EPI_ISL_412973 from the province of Lombardy) isolated in February 2020, grouped separately into a second clade (**Figure 1B Clade 2**) including isolates from Germany, Finland, Switzerland and Mexico. Phylogenetic analyses further suggest that the coronavirus epidemic might have been reintroduced to the country via Germany, since the basal sequence of this clade is a case that emerged in Munich in Germany in January 2020, highlighting a complex dynamic of transmission and a tight connection within European countries.

DISCUSSION

In this study, we investigate the transmission path of the novel SARS-CoV-2 in Italy. Our results indicate that the SARS-CoV-2 circulation in Italy likely started from different multiple introductions from China and Germany (**Figure 1, clade 1 and 2**) (suggesting a possible European linkage) followed by autochthonous transmission within Italy. Moreover, more genomic data will be fundamental in order to shed light on the epidemiological dynamics of the virus in the country helping of understanding regarding how the virus is spreading both around the world and through time. More genomic surveillance efforts during epidemics should be undertaken in order to provide enough data to make reliable inferences on the molecular epidemiology of rapidly evolving virus. Phylogenetic inference robustness will be improved if other genomic sequences will be available to build even more informative dataset despite the low heterogeneity of the sequenced strains. Consequently, this analysis has to be considered

a first preliminary report regarding the SARS-CoV-2 epidemic in Italy that highlights two independent introductions probably mediated by China and Germany respectively. This could reinforce the tight connection within samples that have been isolated all over Europe, indicating that the virus is spreading quite widely in Europe suggesting that it could be likely spreading in the general population.

In conclusion, our study show that genomic data generated in real-time can be employed to assist public health in monitoring and understanding outbreak of novel emerging viral pathogens.

REFERENCES

1. Cui J, Li F, Shi Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;17(3):181–92.
2. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virol J* 2019; 16(1):69.
3. Peiris JS, Guan Y, Yuen KY. Severe acute respiratory syndrome. *Nat Med* 2004;10 (suppl 12): S88–97.
4. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012; 367: 1814–20
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. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020;382(8):727-33.

6. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: Evidence for virus evolution. *J Med Virol* 2020;92:455–59.
7. WHO, Report, March 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_2
8. WHO, Report, February 2020. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200208-sitrep-19-ncov.pdf>
9. Italian Ministry of Health, March, 2020, <http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus.jsp?lingua=italiano&id=5351&area=nuovoCoronavirus&menu=vuoto>
10. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2019;20(4):1160-6.
11. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32(1):268-74.
12. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012;9(8):772.
13. Benvenuto D, Giovanetti M, Salemi M, Prosperi M, De Flora C, Alcantara LC, Angeletti S, and Massimo Ciccozzi. The global spread of 2019-nCoV: a molecular evolutionary analysis. *Pathogen Global Health* 2020;12:1-4

ACKNOWLEDGMENTS

The authors thank all personnel from Health Surveillance Systems who coordinated surveillance in Italy. MG is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ.

FIGURE LEGEND

Figure 1. Maximum likelihood phylogeny, estimated from complete and near-complete CoV genomes using genome data available in GISAID. Colours represent different locations. (Panel A and B represent expansions of the clades containing the Italian CoV isolates (green) in clade 1 closely related with sequence from England, in clade B closely related with sequences from Germany).

