

# Geographical reconstruction of the SARS-CoV-2 outbreak in Lombardy (Italy) during the early phase

Valeria Micheli<sup>1</sup>  | Sara G. Rimoldi<sup>1</sup>  | Francesca Romeri<sup>1</sup> | Francesco Comandatore<sup>2</sup> | Alessandro Mancon<sup>1</sup>  | Anna Gigantiello<sup>1</sup> | Matteo Perini<sup>2</sup>  | Davide Mileto<sup>1</sup> | Cristina Pagani<sup>1</sup> | Alessandra Lombardi<sup>1</sup> | Maria R. Gismondo<sup>1</sup>

<sup>1</sup>Laboratory of Clinical Microbiology, Virology and Bioemergencies, ASST Fatebenefratelli Sacco, L. Sacco University Hospital, Milan, Italy

<sup>2</sup>Department of Biomedical and Clinical Sciences, Romeo ed Enrica Invernizzi Pediatric Research Center, L. Sacco University Hospital, University of Milan, Milan, Italy

## Correspondence

Sara G. Rimoldi, Laboratory of Clinical Microbiology, Virology and Bioemergencies, ASST Fatebenefratelli Sacco, Via G.B. Grassi 74, 20157 Milan, Italy.

Email: [sara.rimoldi@asst-fbf-sacco.it](mailto:sara.rimoldi@asst-fbf-sacco.it)

## Abstract

The first identification of autochthonous transmission of SARS-CoV-2 in Italy was documented by the Laboratory of Clinical Microbiology, Virology and Bioemergencies of L. Sacco Hospital (Milano, Italy) on 20th February 2020 in a 38 years old male patient, who was found positive for pneumonia at the Codogno Hospital. Thereafter Lombardy has reported the highest prevalence of COVID-19 cases in the country, especially in Milano, Brescia and Bergamo provinces. The aim of this study was to assess the potential presence of different viral clusters belonging to the six main provinces involved in Lombardy COVID-19 cases in order to highlight peculiar province-dependent viral characteristics. A phylogenetic analysis was conducted on 20 full length genomes obtained from patients addressing to several Lombard hospitals from February 20th to April 4th, 2020, aligned with 41 Italian viral genome assemblies available on GISAID database as of 30th March, 2020: two main monophyletic clades, containing 8 and 53 isolates, respectively, were identified. Noteworthy, Bergamo isolates mapped inside the small clade harbouring M gene D3G mutation. The molecular clock analysis estimated a cluster divergence approximately one month before the first patient identification, supporting the hypothesis that different SARS-CoV-2 strains had spread worldwide at different times, but their presence became evident only in late February along with Italian epidemic emergence. Therefore, this epidemiological reconstruction suggests that virus initial circulation in Lombardy was ascribable to multiple introduction. The phylogenetic reconstruction robustness, however, will be improved when more genomic sequences are available, in order to guarantee a complete epidemiological surveillance.

## KEYWORDS

genetic variability, pandemics, SARS coronavirus

## 1 | INTRODUCTION

In late December 2019, the Chinese health authorities informed the World Health Organization (WHO) about a pneumonia outbreak of unknown etiology in Wuhan municipality.<sup>1</sup> Thanks to next-generation sequencing approach, a new coronavirus was quickly identified as the causative agent of the new coronavirus disease 2019 (COVID-19)<sup>2</sup>; moreover, due to the rapid global diffusion, WHO declared the pandemic status on 11 March 2020.<sup>3</sup>

The positive-strand RNA virus family Coronaviridae includes more than 40 species, infecting a broad spectrum of vertebrates; in particular, seven coronaviruses are related to human diseases, among which severe acute respiratory syndrome-related coronavirus (SARS-CoV) and the Middle East respiratory syndrome-related coronavirus represented a public health concern in the past 20 years.<sup>4,5</sup> The COVID-19 related virus was classified as a betacoronavirus and, considering its close correlation to SARS-CoV, it was renamed SARS-CoV-2.<sup>6</sup>

In Europe, Italy is one of the most affected areas, accounting for more than 230 000 cases on 5 June 2020.<sup>7</sup> Northern Italy has reported the highest prevalence in the country, especially in Milano, Brescia, and Bergamo provinces, which registered more than 23 000, 14 000, and 13 000 cases,<sup>7</sup> respectively, and an infection rate almost double (0.58% and 0.52%, respectively) compared to the rest of Italy (0.32%). In particular, Bergamo and Brescia provinces had to face a high percentage of severe clinical cases presenting an enormous rate of mortality.

The first identification of autochthonous transmission of SARS-CoV-2 in Italy was documented by the Laboratory of Clinical Microbiology, Virology and Bioemergencies of L. Sacco University Hospital (Milano, Italy) on 20 February 2020: the patient was a 38 years old male, who was found positive for pneumonia at the Codogno Hospital, without any evident linkage with COVID-19 cases; soon after the laboratory received thousands of respiratory samples for the confirmation of suspected COVID-19 from many regional institutions (Bergamo, Brescia, Cremona, Codogno, Lodi, and Milano); owing to the geographical distribution of these specimens, viral sequence data could give insight into SARS-CoV-2 molecular epidemiology and possible local virus introduction.

The aim of the present study was to assess the potential presence of different viral clusters belonging to the six main provinces involved in Lombardy COVID-19 cases to highlight peculiar province-dependent viral characteristics.

## 2 | MATERIALS AND METHODS

### 2.1 | Study population

The study included SARS-CoV-2 positive samples collected at the Laboratory of Clinical Microbiology, Virology and Bioemergencies of L. Sacco University Hospital, a referral center for COVID-19 diagnosis; all specimens were either nasopharyngeal swabs or bronchoalveolar lavage fluid. To optimize sequencing performance, only samples positive for all real-time polymerase chain reaction targets (E gene, N gene,

and ORF1ab; novel coronavirus 2019 Real Time Multiplex RT-PCR Kit, Liferiver) and with a  $C_t \leq 25$  were considered. Since the laboratory received samples from other hospitals widely distributed on the regional territory, patients' location was highly variable; the selection was, therefore, performed aiming at maximizing geographical representation and detection of related diversity. Relevant clinical, demographic, and geographical data were recorded.

### 2.2 | RNA viral extraction and reverse transcription

Total RNA was extracted from 200  $\mu$ L of sample and eluted in 100  $\mu$ L by QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany), according to manufacturer's instructions. Quality, quantity, and purity of the genomic RNA were determined using Qubit 4 Fluorometer (Thermo Fisher Scientific Inc, Monza, Italy). Complementary DNA (cDNA) was synthesized using ImProm-II Reverse Transcriptase (RT) and related reagents (Promega Corporation, Italia). The reaction mixture was prepared as follows: 6  $\mu$ L of ImProm-II 5X Reaction Buffer, 1.8  $\mu$ L of 5 mM  $Mg^{2+}$ , 1  $\mu$ L of 10 mM random primers, 1  $\mu$ L of 5 mM dNTPs, 1  $\mu$ L of RT, 0.5  $\mu$ L of RNasin (40 U/ $\mu$ L), and 20  $\mu$ L of purified RNA, for a total volume of 31.3  $\mu$ L; RT was performed at the following conditions: 37°C for 45 minutes and 80°C for 5 minutes.

### 2.3 | Whole-genome sequencing

cDNA was amplified using two SARS-CoV-2 specific primers sets, to cover the whole viral genome. cDNA was then digested, ligated to barcodes, purified, and amplified again; a second purification was performed before double-stranded DNA quantification on Qubit 4 and library preparation on One Touch 2 Instrument (Thermo Fisher Scientific); the One Touch ES Instrument (Thermo Fisher Scientific) was used for final enrichment and Ion Torrent Personal Genome Machine System (Thermo Fisher Scientific) for sequencing, following manufacturer's instructions.

### 2.4 | Genome assembly

For each sample, the genome assembly was obtained using a mapping-based approach. Low-quality reads bases were trimmed out using Trimmomatic software,<sup>8</sup> using 13 different parameter sets (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36, LEADING:3 TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:36, LEADING:3 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:36, LEADING:3 TRAILING:10 SLIDINGWINDOW:4:25 MINLEN:36, LEADING:3 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:36, LEADING:3 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36, LEADING:3 TRAILING:20 SLIDINGWINDOW:4:25 MINLEN:36, MAXINFO:50:0.3, MAXINFO:50:0.5,

MAXINFO:50:0.7, MAXINFO:50:0.9). Then, single-nucleotide polymorphism (SNP) calling was performed following the GATK Best Practice procedure,<sup>9</sup> using the Wuhan-Hu-1 strain genome (accession MN908947.3) as reference. The genome consensus sequence was obtained on the basis of the identified SNPs and the reference sequence. Reference bases were called in conserved positions with coverage above 5, otherwise, N was introduced. The procedure produced 13 genome assemblies per sample, corresponding to the 13 different trimming parameter sets. For each sample, the genome assembly with the lowest number of undetermined bases (N) was selected as the final genome assembly.

## 2.5 | Molecular clock analysis

A dataset of 41 genome assemblies of SARS-CoV-2 strains isolated in Italy between 20 February 2020 and 30 March 2020 was retrieved from Global Initiative on Sharing All Influenza Data (GISAID) database (Table S1). A global dataset including these 41 GISAID genome

assemblies and the 20 genome assemblies generated in this study was produced and aligned using MAFFT.<sup>10</sup>

The low-quality alignment regions at the extremities of the alignment were removed using Gblocks with default parameters.<sup>11</sup> The alignment was subjected to maximum likelihood phylogenetic analysis using RaxML.<sup>12</sup> The obtained tree was then analyzed using TempEst<sup>13</sup> to assess temporal signals. The Hasegawa-Kishino-Yano model was found as the simplest evolutionary model using JmodelTest 2.1.10.<sup>14</sup> Phylogenetic analysis was performed using a Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST, v.1.10.4<sup>15</sup> with 10 million states and sampling every 1000 steps. The coalescent priors constant population size and exponential growth were tested for a relaxed molecular clock using path sampling (PS) and stepping stone (SS) sampling.<sup>16</sup> Then, the phylogenetic analysis was repeated with the selected coalescent prior with 100 million states and sampling every 1000 steps. The convergence of the MCMC chain was checked using Tracer v.1.7.1. The maximum clade credibility trees were obtained from the tree posterior distribution using TreeAnnotator (<http://beast.community/index.html>).

**TABLE 1** Dataset of 20 genome assemblies of SARS-CoV-2 strains from patients addressing to several Lombard hospitals from 20 February to 4 April 2020 tested at Sacco Hospital

ID	GISAID virus name	City	Province	Hospital	Ward/hospital	Collection date
HSacco-2	hCoV19/Italy/Hsacco-2/2020	n/a	n/a	Hospital Luigi Sacco	MI2	2020-03-09
HSacco-3	hCoV19/Italy/Hsacco-3/2020	Lodi	Lodi	ASST Lodi	n/a	2020-02-22
HSacco-4	hCoV19/Italy/Hsacco-4/2020	Lodi	Lodi	ASST Lodi	n/a	2020-02-23
HSacco-5	hCoV19/Italy/Hsacco-5/2020	Bergamo	Bergamo	Nursing home S. Francesco	n/a	2020-03-02
HSacco-6	hCoV19/Italy/Hsacco-6/2020	Bergamo	Bergamo	Nursing home S. Francesco	n/a	2020-03-02
HSacco-7	hCoV19/Italy/Hsacco-7/2020	Locate di Triulzi	Milano	Hospital Luigi Sacco	MI1	2020-03-06
HSacco-8	hCoV19/Italy/Hsacco-8/2020	Lodi	Lodi	Hospital Luigi Sacco	Intensive Care Unit	2020-02-26
HSacco-9	hCoV19/Italy/Hsacco-9/2020	Castrezzato	Brescia	ASST Franciacorta	n/a	2020-03-03
HSacco-10	hCoV19/Italy/Hsacco-10/2020	Azzano Mella	Brescia	ASST Franciacorta	n/a	2020-03-02
HSacco-11	hCoV19/Italy/Hsacco-11/2020	Villolongo	Bergamo	n/a	n/a	2020-03-02
HSacco-12	hCoV19/Italy/Hsacco-12/2020	Milano	Milano	Nursing home Igea	n/a	2020-03-13
HSacco-13	hCoV19/Italy/Hsacco-13/2020	Crema	Crema	Hospital Luigi Sacco	Intensive Care Unit	2020-03-17
HSacco-14	hCoV19/Italy/Hsacco-14/2020	Milano	Milano	Hospital Luigi Sacco	Oncology	2020-03-24
HSacco-15	hCoV19/Italy/Hsacco-15/2020	Milano	Milano	Hospital Luigi Sacco	Oncology	2020-04-02
HSacco-16	hCoV19/Italy/Hsacco-16/2020	Milano	Milano	Hospital Luigi Sacco	Oncology	2020-03-20
HSacco-17	hCoV19/Italy/Hsacco-17/2020	n/a	n/a	n/a	n/a	2020-04-04
HSacco-18	hCoV19/Italy/Hsacco-18/2020	Milano	Milano	n/a	n/a	2020-04-02
HSacco-19	hCoV19/Italy/Hsacco-19/2020	Codogno	Lodi	n/a	n/a	2020-02-20
HSacco-20	hCoV19/Italy/Hsacco-20/2020	Bergamo/Lodi	Bergamo/Lodi	n/a	n/a	2020-02-26
HSacco-21	hCoV19/Italy/Hsacco-21/2020	n/a	n/a	n/a	n/a	2020-04-01

Note: Milano province includes also patients not living in Milan, who were already hospitalized at Sacco Hospital, suggesting a nosocomial acquisition. Abbreviations: GISAID, Global Initiative on Sharing All Influenza Data; n/a, not applicable; SARS-CoV-2, severe acute respiratory syndrome-related coronavirus 2.

## 2.6 | Ethical approval

All data used in the study were previously made anonymous, according to the requirements set by Italian Data Protection Code (Legislative Decree 196/2003) and the general authorizations issued by the Data Protection Authority. Under Italian law, all sensitive data were deleted and only age, sex, and sampling date were collected making Ethics Committee approval unnecessary (Art. 6 and Art. 9 of Legislative Decree 211/2003).

## 3 | RESULTS

### 3.1 | Study population

A total of 20 samples were sequenced and included in phylogenetic analysis, attributing the progressive ID HSacco-N (from HSacco-2 to HSacco-21). All patients were resident in Lombardy, distributed in different provinces; in particular, the province "Milano" contains also patients hospitalized for non-COVID-19 disease, for whom SARS-CoV-2 hospital acquisition was supposed; in addition, HSacco-20 was a nurse living in Bergamo and working at Lodi Hospital.

### 3.2 | Phylogenetic and molecular clock analysis

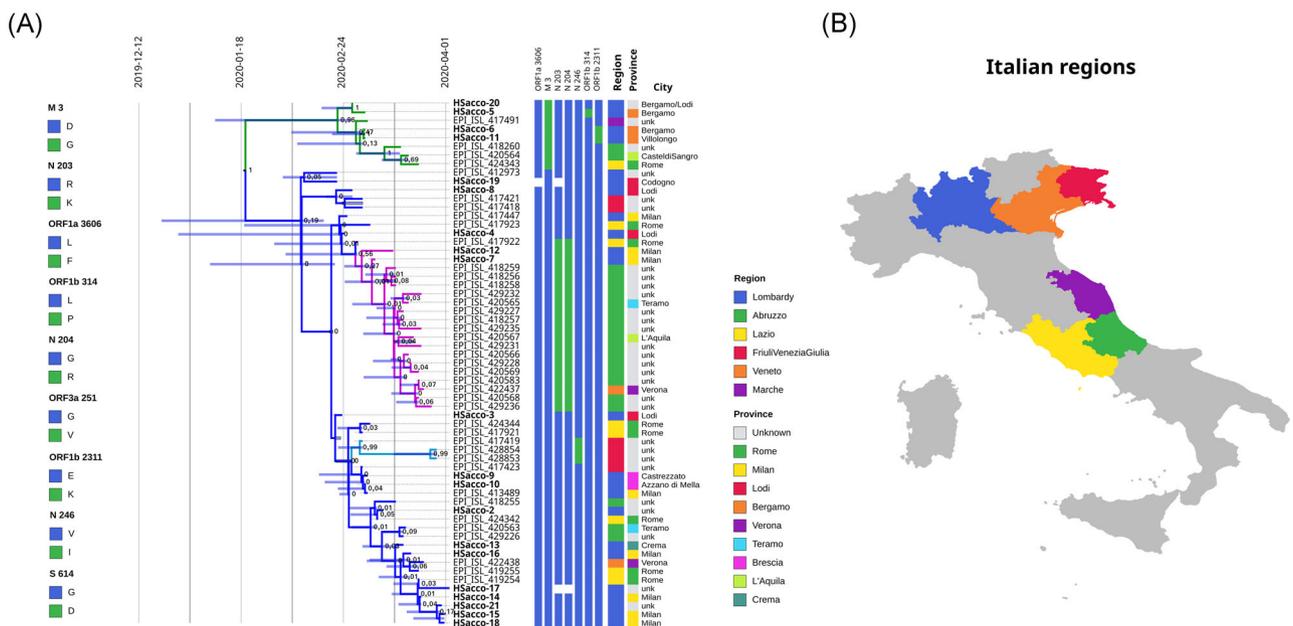
The 20 SARS-CoV-2 genome assemblies are available on the GISAD database (Table 1). The comparison of the marginal likelihoods of constant and exponential coalescent models under a log-normal

relaxed clock showed that the best fitting model was the exponential coalescent prior (PS Bayes factors [BF] exponential growth vs constant = 2,42; SS BF exponential growth vs constant = 2,39). The obtained phylogenetic tree is reported in Figure 1.

The tree shows the existence of two major monophyletic clades, containing 8 and 53 isolates, respectively (in green and blue, respectively, in Figure 1). The smaller clade is characterized by a nonsynonymous mutation in position three of M gene (mutation D3G), which time of the most recent common ancestor (tMRCA) resulted 20 February 2020 (highest posterior density [HPD] 95% interval: 4 January to 24 February); Bergamo and HSacco-20 sequences mapped inside this group. The larger clade contains two monophyletic subclades, characterized by nonsynonymous mutations in the N gene (Figure 1). The first cluster includes 19 isolates, mainly from Central Italy, in which R203K and G204R variants were found; the other one, instead, accounted for three sequences from Friuli Venezia Giulia, presenting the V246I substitution. The tMRCA calculated for the two subclades were 2 March 2020 (HPD 95% interval: 24 February to 6 March) and 28 February 2020 (HPD 95% interval: 20 February to 2 March), respectively (Figure 1).

## 4 | DISCUSSION

Lombardy had the highest prevalence of COVID-19 in Italy, thus being the likely epicenter of country outbreak.<sup>7</sup> However, it is still unclear how SARS-CoV-2 circulation started: no epidemiological link was found for the first identified autochthonous patient. Moreover, emerging evidence suggest a multiple virus introduction at least in January 2020.<sup>17,18</sup>



**FIGURE 1** Molecular clock, single-nucleotide polymorphisms (SNPs), and geographic information. A, In the figure, information from tree dating, SNPs, and geographic origin are reported. On the left, the dated tree with the posterior probabilities and the highest posterior density 95% bars reported on the tree nodes; on the right, SNPs and geographic information are shown using the color code reported in the legends. B, Geographic map of Italy, with the regions from which isolates were collected, colored following the color code of the "Region" legend in (A)

This study tried to expand understanding of SARS-CoV-2 circulation during the early epidemic period in Lombardy, especially in the most affected areas. A phylogenetic analysis was conducted on 20 full-length genomes from patients addressing to several Lombard hospitals (20 February to 4 April 2020), aligned with 41 Italian viral genome assemblies from GISAID database as of 30 March 2020. In spite of restricted geographical origin, these sequences were included in two main monophyletic clades containing isolates collected across different regions; in addition, the main clade of 53 assemblies was divided into two subclades. Interestingly, the molecular clock analysis estimated a cluster divergence approximately 1 month before the first patient identification. Such evidence are in accordance with other studies, further supporting the hypothesis that different SARS-CoV-2 strains had spread worldwide at different times, but their presence became evident only in late February along with Italian epidemic emergence.<sup>17,19</sup>

Noteworthy, Bergamo (HSacco-5, HSacco-6, and HSacco-11) and HSacco-20 sequences had M gene D3G mutation of the first clade, in contrast with all the other ones located in the most represented clade. Two main observations arose: the virus found in Bergamo area seemed to have a more restricted circulation, probably for a delayed introduction. In addition, the nurse working in Lodi and living in Bergamo likely acquired the infection in a personal contest rather than the hospital environment, since Lodi patients mapped inside the other cluster.

Another interesting outcome was the presence of N gene mutations R203K and G204R subclade. These amino-acid changes appear to coexist, being always found together in the dataset, as well as present almost exclusively in GISAID 20B clade. The phylogenetic tree also clearly showed that the subclade is predominantly made up of isolates from Abruzzo, suggesting segregation of this specific virus in the Central Italy area. Since the role of these mutations is still unclear, an Indian group investigated their possible influence on virus replication interference mediated by microRNA (miRNA): they found out how some miRNAs, present in different pathological conditions, are likely to bind to native N gene and repress its expression, thus helping in disease progression limitation; on the contrary, mutated variants could increase their chances of interference escape.<sup>20</sup>

Another small subclade was found in the main one, characterized by N gene V246I mutation (three sequences from Friuli Venezia Giulia). Besides the unique geographical origin, it is noticeable that in GISAID map "Geography" the V246I mutation is actually present only in Italy, as well as the V246A one only in Israel. Their rarity could have two probable explanations: on the one hand, available data are still limited, making it difficult to have a reliable distribution; on the other hand, these variations could have a negative influence on viral fitness, diminishing virus replication and/or transmission efficacy.

Other mutations found in this study had negligible influence on phylogenetic analysis, even if a biological significance cannot be excluded: viral genome and proteins are key factors in patients' management and any variation can extremely impact drugs, vaccines, and diagnostic tools or be related to a more severe clinical presentation.

In conclusion, this study gave insights on early dynamics of SARS-CoV-2 circulation in Italy, underlying peculiar strains localization and supporting multiple virus introductions at least in January 2020.

## ACKNOWLEDGMENTS

The authors would like to thank Gherard Batisti Biffignandi for graphic support and all the Members of Laboratory of Clinical Microbiology, Virology and Bioemergencies as follows: Rizzo Alberto, Giacomel Giovanni, Curreli Daniele, Cappelletti Cristina, Zanchetta Nadia, Grosso Silvia Nella Faustina, Longo Margherita Olga Saveria, Lanzafame Salvatore Maria Giovanni, Longobardi Concetta, Calvagna Annunziata, Bosari Raffaella, Maresca Mafalda, Caliano Rosanna, Banfi Daniela, Tonielli Claudia, Bossi Carla, Fiori Lorenza, Tamoni Alessandro, Secchi Daniela, Pontoriero Cecilia, Dichirico Rita Barbara, Bertulli Manuela, Legnani Sara, Loiacono Aurelia, Terron Enrica, Di Paola Agata, and Razzitti Vittoria Paola.

## DATA AVAILABILITY STATEMENT

Sequences were submitted to GISAID.

## ORCID

Valeria Micheli  <http://orcid.org/0000-0001-5629-3875>

Sara G. Rimoldi  <http://orcid.org/0000-0003-1935-8750>

Alessandro Mancon  <https://orcid.org/0000-0002-1490-636X>

Matteo Perini  <http://orcid.org/0000-0002-9465-6216>

## REFERENCES

- World Health Organization. *Novel coronavirus (2019-nCoV), situation report-1*, 21 January 2020. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10\\_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4). Accessed January 30, 2020.
- Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273.
- World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020*. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020>. Accessed March 20, 2020.
- Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-574.
- De Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol*. 2016;14:523-534.
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5:536-544.
- Dipartimento della Protezione Civile. *COVID-19 Italia—Monitoraggio della situazione*. <http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1>. Accessed June 5, 2020.
- Bolger A, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
- DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-498.

10. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*. 2018; 34(14):2490-2492.
11. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000; 17(4):540-552.
12. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9): 1312-1313.
13. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst. *Virus Evol*. 2016;2(1):vew007.
14. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9(8):772.
15. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018;4(1):vey016.
16. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. 2012;29(9):2157-2167.
17. Giovanetti M, Angeletti S, Benvenuto D, Ciccozzi M. A doubt of multiple introduction of SARS-CoV-2 in Italy: a preliminary overview. *J Med Virol*. 2020;92:1634-1636.
18. Zehender G, Lai A, Bergna A, et al. Genomic characterization and phylogenetic analysis of SARS-CoV-2 in Italy. *J Med Virol*. 2020;92:1637-1640.
19. van Dorp L, Acman M, Richard D, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol*. 2020;83:104351.
20. Maitra A, Sarkar MC, Raheja H, et al. Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. *J Biosci*. 2020;45:76.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Micheli V, Rimoldi SG, Romeri F, et al. Geographical reconstruction of the SARS-CoV-2 outbreak in Lombardy (Italy) during the early phase. *J Med Virol*. 2021;93: 1752-1757. <https://doi.org/10.1002/jmv.26447>