

The first two cases of 2019-nCoV in Italy: Where they come from?

Marta Giovanetti¹ | Domenico Benvenuto²  | Silvia Angeletti³  | Massimo Ciccozzi² 

¹Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

²Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy

³Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome, Italy

Correspondence

Silvia Angeletti, Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome 00128, Italy.

Email: s.angeletti@unicampus.it

Abstract

A novel *Coronavirus*, 2019-nCoV, has been identified as the causal pathogen of an ongoing epidemic, with the first cases reported in Wuhan, China, last December 2019, and has since spread to other countries worldwide, included Europe and very recently Italy. In this short report, phylogenetic reconstruction was used to better understand the transmission dynamics of the virus from its first introduction in China focusing on the more recent evidence of infection in a couple of Chinese tourists arrived in Italy on 23rd January 2020 and labeled as *Coronavirus* Italian cases. A maximum clade credibility tree has been built using a dataset of 54 genome sequences of 2019-nCoV plus two closely related bat strains (SARS-like CoV) available in GenBank. Bayesian time-scaled phylogenetic analysis was implemented in BEAST 1.10.4. The Bayesian phylogenetic reconstruction showed that 2019-2020 nCoV firstly introduced in Wuhan on 25 November 2019, started epidemic transmission reaching many countries worldwide, including Europe and Italy where the two strains isolated dated back 19 January 2020, the same that the Chinese tourists arrived in Italy. Strains isolated outside China were intermixed with strains isolated in China as evidence of likely imported cases in Rome, Italy, and Europe, as well. In conclusion, this report suggests that further spread of 2019-nCoV epidemic was supported by human mobility and that quarantine of suspected or diagnosed cases is useful to prevent further transmission. Viral genome phylogenetic analysis represents a useful tool for the evaluation of transmission dynamics and preventive action.

KEYWORDS

epidemiology, engineering and technology, infection, macromolecular design

1 | INTRODUCTION

An ongoing epidemic by a new *Coronavirus*, named 2019-nCoV, starting in late 2019 in Wuhan, Hubei region, China, is a worldwide public health concern.¹⁻³ The virus probably originated from bat after mutation in the spike glycoprotein, as recently suggested, acquired the ability to infect humans, which started the new epidemic.⁴ As of today (31 January 2020), 14 628 total confirmed cases have been reported, with 14 451 cases in China and the remaining cases distributed among countries in every continent, but predominantly in

Japan, Thailand, Singapore, South Korea, Hong Kong, Australia, and Taiwan. In Europe, the first few cases have been reported in Germany and France and most recently in Italy and Spain. The death toll from the 2019-nCoV outbreak is now 305 and the number of totals recovered 348. (<https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>). In Italy, two cases have been registered in Rome, in a couple of Chinese tourists hosted in a Hotel placed in the center of the city. The Chinese tourists had arrived in Milan on 23 January 2020 from Wuhan, and thereafter they traveled to Rome. After relevant

symptoms of coronavirus diseases, the couple has been admitted to Rome's Spallanzani Hospital, which is specialized in infectious diseases. In Italy, a state of emergency has been declared as a consequence of these two confirmed cases of infection that have been labeled as the first cases of *Coronavirus* transmission described in Italy.

In this report, phylogenetic and evolutionary analysis has been applied to characterize the 2019-nCoV virus identified in Italy and better understand the transmission dynamics of the two cases diagnosed in Rome.

2 | MATERIALS AND METHODS

The dataset used for phylogenetic analysis included ($n = 54$) genomes sequences from the current 2019-nCoV epidemic plus ($n = 2$) closely related bat strains (SARS-like CoV) retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/genbank/>) and GISAID (<https://www.gisaid.org/>) databases. Sequences were aligned using MAFFT software⁵ and manually edited by BioEdit program v7.0.5.⁶

By IQ-TREE 1.6.8 software the maximum likelihood (ML) phylogeny was reconstructed under the HKY nucleotide substitution model with 4 gamma categories (HKY+G4) which were inferred in jModelTest (<https://github.com/ddarriba/jmodeltest2>) as the best fitting model.⁷

To investigate the temporal signal, from the ML tree, we regressed root-to-tip genetic distances against sample collection dates using TempEst v 1.5.1 (<http://tree.bio.ed.ac.uk>).⁸ Bayesian time-scaled phylogenetic analysis was performed by BEAST 1.10.4 (<http://beast.community/index.html>), using ML phylogeny as starting tree.⁹ A stringent model selection analysis by both path sampling (PS) and stepping stone (SS) procedures were performed to estimate the most appropriate molecular clock model for the Bayesian phylogenetic analysis.¹⁰ The strict molecular clock model, assuming a single rate across all phylogeny branches, and the more flexible uncorrelated relaxed molecular clock model with a lognormal rate distribution were tested.¹¹ Both SS and PS estimators indicated the uncorrelated relaxed molecular clock (Bayes factor = 4.3) as the best-fitted model to the dataset under analysis. The HKY+G4 codon partitioned 1 + 2,3 substitution model and the Bayesian skyline coalescent model of population size and growth was implemented.¹² Markov chain Monte Carlo (MCMC) duplicate runs of 100 million states each, sampling every 10 000 steps was computed. The convergence of MCMC chains was checked using Tracer v.1.7.1.¹² Proper mixing of the MCMC was checked for ESS values greater than 200 for each estimated parameter using Tracer 1.7. *Syst Biol.* 2018;67(5):901-904. The maximum clade credibility (MCC) trees were obtained from the tree posterior distribution using TreeAnnotator (<http://beast.community/index.html>) after 10% burn-in.

3 | RESULTS

The viral genomes analyzed from the dataset, despite being isolated in a short time, already exhibited a substantial degree of

heterogeneity with differences in 15% of the sites, 11% of which were parsimony informative. These data are coherent with the presence of sufficient phylogenetic signal for phylogenetic inference, in agreement with the low level of phylogenetic noise shown by likelihood mapping (<7%). The root-to-tip vs divergence plot of the full dataset showed high correlation between sampling time and genetic distance to the root of the ML tree of the available sequences ($R^2, .60$), coherently with substantial temporal signal and the possibility to calibrate a reliable molecular clock, despite availability of relative small number of sequences and short sampling interval time.

Bayesian model selection chose the Bayesian skyline demographic model with an uncorrelated relaxed clock like the one best fitting the data. Molecular clock calibration estimated the evolutionary rate of the 2019-nCoV whole-genome sequences at 6.58×10^{-3} substitutions site per year (95% HPD 5.2×10^{-3} to 8.1×10^{-3}).

In Figure 1, the MCC tree with Bayesian phylogenetic reconstruction of 2019-nCoV isolates up to day is shown. The probable origin of 2019-nCoV was confirmed to be Wuhan with a state posterior probability (spp) of 0.93 datings back to the time of the most recent common ancestor of the human outbreak reported on 25th November 2019 (95% HPD, 28 September 2019; 21 December 2019). The dissemination of 2019-nCoV, according to with the phylogenetic reconstruction, is confirmed to be from Wuhan, China, to Guangdong, Zhuhai (spp, 0.90), to Nonthaburi, Thailand, with an spp. of 0.90. From China, our reconstruction revealed independent importation events in the USA including Washington, Chicago, Illinois, Arizona, and Los Angeles, in Australia and in Europe including France, Germany, and Italy (spp 0.90). The viral sequence from the two Chinese tourist patients, who received a diagnosis of *Coronavirus* infection after their arrival in Italy on 23 January 2020 and during their stay in Rome, clustered together and grouped with sequences isolated in Europe intermixing with viral strains from China. The strains from the Chinese tourist dated back to 19 January 2020 (95% HPD: 18 January 2020; "1 January 2020"; Figure 1).

4 | DISCUSSION

The early sharing of data about the Coronavirus 2019-nCoV epidemic is fundamental to better understand the transmission dynamics of this virus.¹³ This short report is focused on the first cases of 2019-nCoV infection diagnosed in two Chinese tourists upon their arrival in Italy. Bayesian phylogenetic reconstruction suggested that Chinese tourists were probably infected before their arrival in Italy, since the viral strains sequences intermixed with Chinese isolates of the epidemic dating back probably to 19 January 2020 before their arrival in Italy. Furthermore, the Chinese tourist's viral strains isolated in Italy clustered with other European strains from France, and Germany intermixing with other Chinese sequences, thus suggesting that strains introduced in Europe up to day are from China. While the

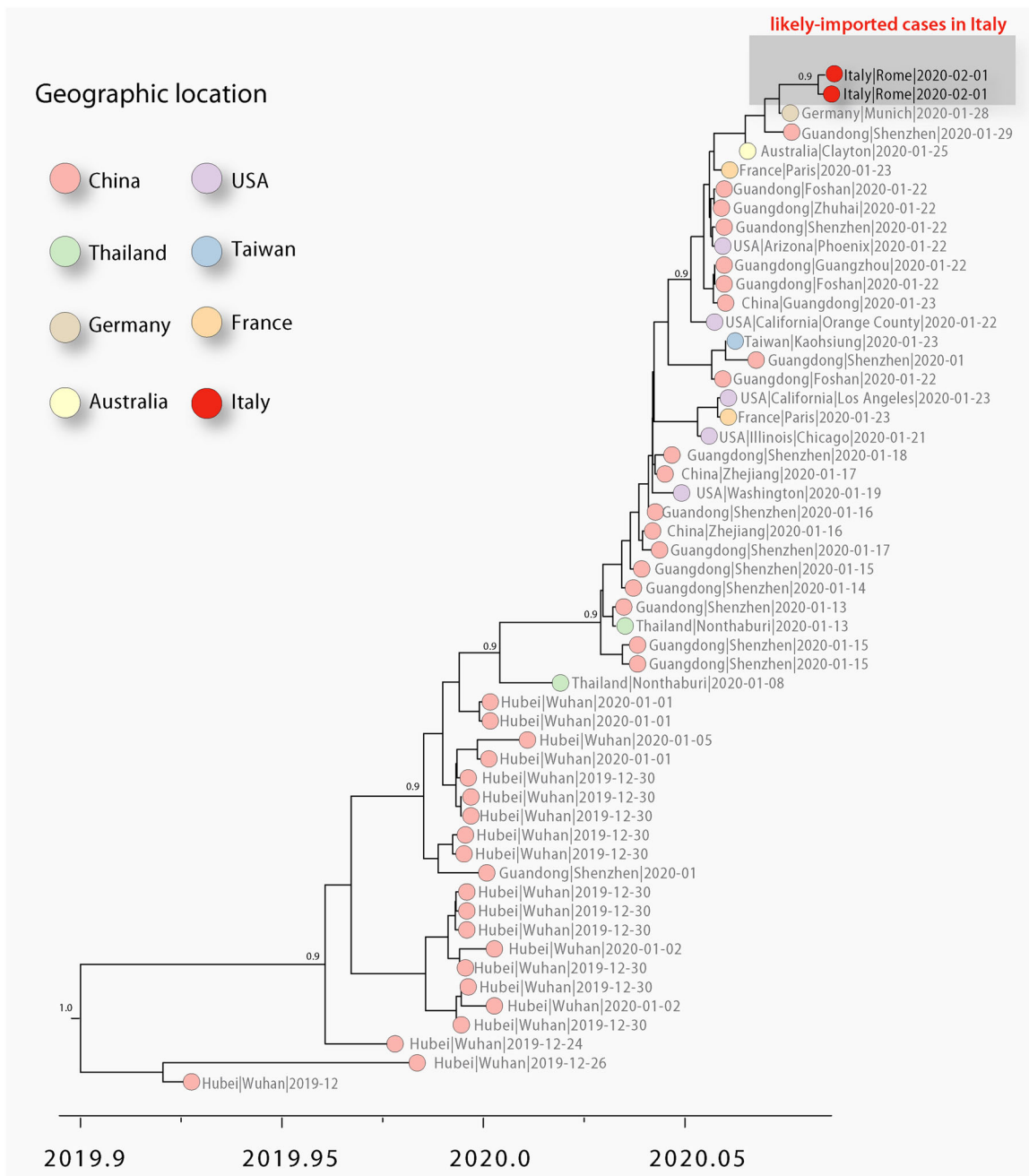


FIGURE 1 Maximum clade credibility phylogeny, estimated from complete and near-complete nCoV virus genomes with a molecular clock phylogenetic approach. Expansion of the clade containing the novel genomes sequences from the nCoV 2019–2020 epidemic. Clade posterior probabilities are shown at well-supported nodes. Colors represent different locations

outbreak seems to be centered in Wuhan, China, the genomic analysis and the phylogenetic reconstruction suggested that Wuhan, acting as a source for 2019-nCoV, or as stepping stone spot, favored viral dissemination to other areas in and outside the country, which might have been influenced by the increased human mobility. Here we showed evidence of likely imported cases from China in Rome, Italy, confirming that tourist mobility was responsible for virus isolation in Italy. In conclusion, our study shows that genomic data

generated in real-time can be employed to assist public health laboratories in monitoring and understanding the diversity of emerging viruses. In the globalization era characterized by high human mobility, this report suggests that the spread of a new epidemic is favored, rapidly reaching other countries even in other continents as we observed during the last days. Immediate quarantine and prompt action for infection control represent the best tools to avoid further spread of infection, as the Italian government provided.

ORCID

Domenico Benvenuto  <http://orcid.org/0000-0003-3833-2927>

Silvia Angeletti  <http://orcid.org/0000-0002-7393-8732>

Massimo Ciccozzi  <http://orcid.org/0000-0003-3866-9239>

REFERENCES

- Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020. <https://doi.org/10.1056/NEJMoa2001017>
- Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- Hui DS, I Azhar E, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis*. 2020;91:264-266.
- 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. <https://doi.org/10.1002/jmv.25688>
- Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20(4):1160-1166. <https://doi.org/10.1093/bib/bbx108>
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41:95-98.
- Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9(8):772.
- Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016;2(1):vew007.
- 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.
- Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol Biol Evol*. 2013;30(2):239-243.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018;67(5):901-904.
- The Lancet. Emerging understandings of 2019-nCoV. *Lancet*. 2020;395(10221):311. [https://doi.org/10.1016/S0140-6736\(20\)30186-0](https://doi.org/10.1016/S0140-6736(20)30186-0)

How to cite this article: Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019-nCoV in Italy: Where they come from? *J Med Virol*. 2020;92:518–521. <https://doi.org/10.1002/jmv.25699>