

SARS-CoV-2 Lineages and Sub-Lineages Circulating Worldwide: A Dynamic Overview

Eleonora Cella^a Francesca Benedetti^b Silvia Fabris^c Alessandra Borsetti^d
Aldo Pezzuto^e Marco Ciotti^f Stefano Pascarella^g Giancarlo Ceccarelli^h
Davide Zella^b Massimo Ciccozzi^c Marta Giovanetti^{i,j}

^aBurnett School of Biomedical Sciences, University of Central Florida, Orlando, FL, USA; ^bInstitute of Human Virology, Department of Biochemistry and Molecular Biology, School of Medicine, University of Maryland, Baltimore, MD, USA; ^cMedical Statistic and Molecular Epidemiology Unit, University of Biomedical Campus, Rome, Italy; ^dNational HIV/AIDS Research Center, Istituto Superiore di Sanità, Rome, Italy; ^eCardiovascular-Respiratory Sciences Department, Sant' Andrea Hospital-Sapienza Università, Rome, Italy; ^fDipartimento di Scienze Biochimiche e Centro di Biologia Molecolare del Consiglio Nazionale delle Ricerche, Università La Sapienza, Roma, Italy; ^gDepartment of Biochemical Sciences "A. Rossi Fanelli," University of Rome "La Sapienza," Rome, Italy; ^hDipartimento di Salute Pubblica e Malattie Infettive, Policlinico Umberto I Università "Sapienza," Rome, Italy; ⁱLaboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; ^jLaboratório de Genética Celular e Molecular, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Keywords

SARS-CoV-2 · Lineages · Pandemic · Genome diversity · Prevention strategies

Abstract

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) originated in Wuhan, China, in early December 2019 has rapidly widespread worldwide, becoming one of the major global public health issues of the last centuries. **Key Messages:** Over the course of the pandemic, due to the advanced whole-genome sequencing technologies, an unprecedented amount of genomes have been generated, providing invaluable insights into the ongoing evolution and epidemiology of the virus during the pandemic.

Therefore, this large amount of data played an important role in the SARS-CoV-2 mitigation and control strategies. **Key Messages:** The active monitoring and characterization of the SARS-CoV-2 lineages circulating worldwide is useful for a more specific diagnosis, better care, and timely treatment. In this review, a concise characterization of all the lineages and sub-lineages circulating and co-circulating across the world has been presented in order to determine the magnitude of the SARS-CoV-2 threat and to better understand the virus genetic diversity and its dispersion dynamics.

© 2021 S. Karger AG, Basel

Marta Giovanetti
Reference Laboratory of Flavivirus, Oswaldo Cruz Institute
Ave. Brasil 4365
Rio de Janeiro 21040-360 (Brazil)
giovanetti.marta@gmail.com

Eleonora Cella
Burnett School of Biomedical Sciences, University of Central Florida
4110 Libra Dr.
Orlando, FL 32816 (USA)
eleonora.cella@yahoo.it

Introduction

On December 31, 2019, the WHO country office in China acknowledged unusual viral pneumonia cases in Wuhan. The responsible agent has been identified as a new coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1] and the associated symptoms are referred to as coronavirus disease 2019 (COVID-19) [2]. This virus belongs to the *Betacoronavirus* genus, and its genome is composed of 29,903 nucleotides [3]. Its genome organization is shared with other Betacoronaviruses, and it presents 6 functional open reading frames (ORFs) that are arranged in order from 5' to 3': replicase (ORF1a/ORF1b), spike (S), envelope, membrane, and nucleocapsid [4].

The epidemic peak in China was reached in early February 2020, and the Chinese government implemented strict measures to contain SARS-CoV-2 spread, including gathering restrictions and travel ban [5]. Whereas in China, new cases decreased, worldwide SARS-CoV-2 infections started to increase. Following the viral spread worldwide, the WHO declared the outbreak as a Public Health Emergency of International Concern on January 30, 2020, and a pandemic on March 11, 2020. Since the first detec-

tion in China in December 2019, more than 77 million cases have been registered worldwide with >1.7 million deaths in 218 affected countries. Among these, the United States is the country with the highest number of cases worldwide (16.5 million), followed by India with 9.8 million (updated on December 20, 2020) [6].

To contrast the SARS-CoV-2 pandemic, several countries implemented strict preventive measures such as lockdown, store closure, and restrictive international flight travels that appear to have dictated the profile of the SARS-CoV-2 spread on a local scale [7–10]. Moreover, continued and uncontrolled SARS-CoV-2 transmissions worldwide have also created the conditions for significant virus evolution, allowing for the rapid spread and adaptation of new SARS-CoV-2 variants [11–15]. For this reason, characterization of new variants could lead to a better understanding of the virus transmissibility rate and infection severity and have implications for successful vaccine deployment. In this review, we present a concise characterization of all the lineages and sub-lineages circulating and co-circulating across the world with the aim of determining the magnitude of SARS-CoV-2 threat and better understanding the virus genetic diversity and its dispersion dynamics.

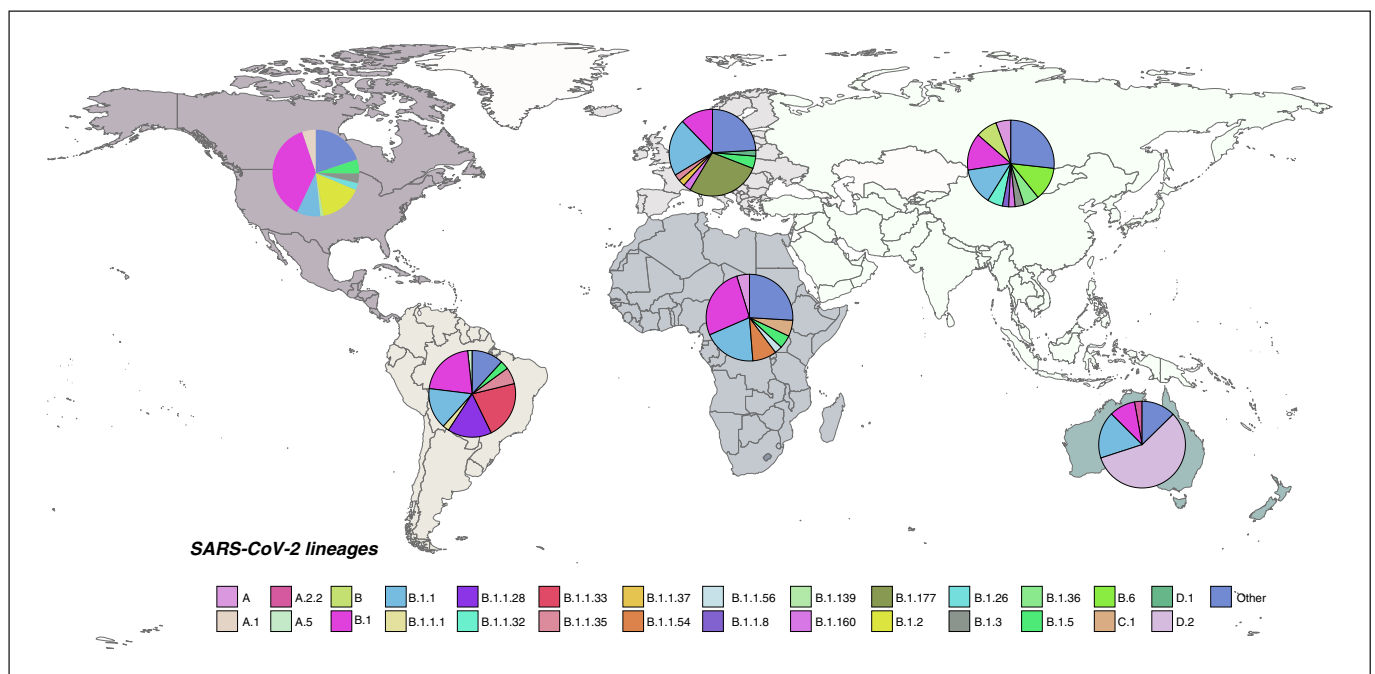


Fig. 1. Worldwide lineage distribution. Lineage with a prevalence percentage <2% were grouped as “other.” (For details of the grouped lineages, see online suppl. Table 1.) SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Main Text

As of today, December 20, 2020, 261,487 complete or near-complete genome sequences are publicly available. One of the first available genome was Wuhan-Hu-1 (MN908947.3) sampled on December 26, 2019 in Wuhan, China [3], obtained only 6 days after the first SARS-CoV-2 notified case. Since there, open and rapid data sharing through the *Global Initiative on Sharing Avian Influenza Data* (GISAID) has enabled evolutionary analysis of SARS-CoV-2 on a large scale and in near real time [16]. This initiative played an important role in global and European pandemic preparedness and increased our understanding on how fast the virus was spreading and evolving across the world [16].

In July 2020, Rambaut et al. [17] designed a dynamic nomenclature for SARS-CoV-2 lineage assignment in order to facilitate the real-time genomic epidemiol-

ogy. By providing commonly agreed labels to refer to viruses circulating in different parts of the world, it was then possible to outline the links between outbreaks that share similar viral genomes. For this purpose, an algorithm named *Phylogenetic Assignment of Named Global Outbreak LINEages* (pangolin) [17, 18], available at <https://github.com/hCoV-2019/pangolin>, was implemented.

According to this classification, 2 major lineages at the root of the phylogeny of the SARS-CoV-2 can be distinguished worldwide, namely, lineages A and B [17]. Lineage A can be defined by the Wuhan/WH04/2020 sequence and appear to share 2 nucleotides (positions 8,782 in ORF1ab and 28,144 in ORF8) with the closest known bat viruses (RaTG13 and RmYN02). While different nucleotides are present at those sites in viruses assigned to lineage B, which is represented by the Wuhan-Hu-1 strain [17].

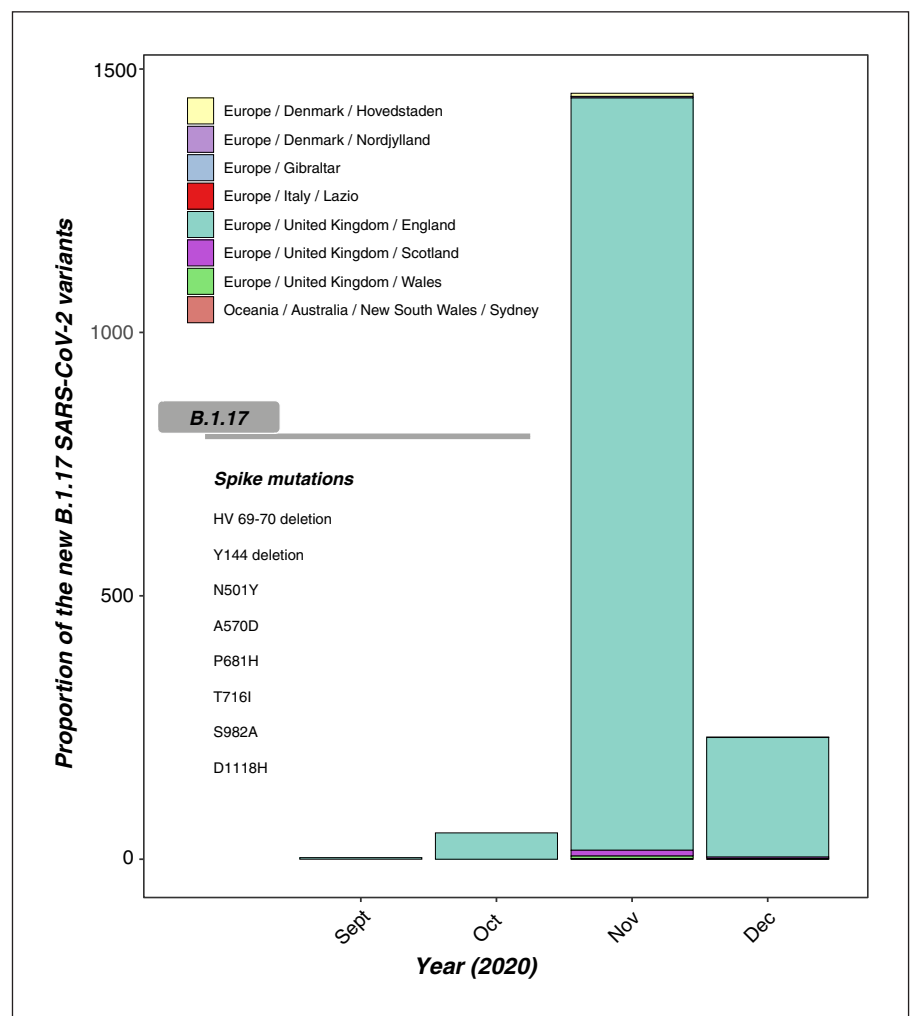


Fig. 2. Proportion of the new B.1.17 SARS-CoV-2 variants among time. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Most likely, the most recent common ancestor (MRCA) of the SARS-CoV-2 phylogeny shares the same genome sequence as the early lineage A sequences (e.g., Wuhan/WH04/2020) [17] even if lineage B was first detected. After the worldwide spread, lineages A and B were further divided into sub-lineages, and other 2 main lineages were detected, namely, lineages C and D that have been recently reassigned as alias of lineage B [12]. Furthermore, about 266 lineages/sub-lineages were identified worldwide (see online suppl. Fig. 1; for all online suppl. material, see www.karger.com/doi/10.1159/000515340) (updated up to December 20, 2020).

Currently, the lineage B and sub-lineage B.1 appear to be the most prevalent worldwide, while in Oceania, the most prevalent lineage appears to be D.2 (Fig. 1; online suppl. Fig. 1). Additionally, lineage B is the most prevalent in African countries (26.7%), B.1.1 in Asia (13.8%), B.1.177 in Europe (27.7%), B.1 in North America (37.8%), and B.1.1.3 in South America (21.6%). Lineages C and D appear to be likely distributed in a local scale. C.1 and C.2 are mainly found in South Africa (99 and 92%, respectively), whereas C.3 and D1 are mainly found in the UK.

The ongoing uncontrolled spreading of the SARS-CoV-2 worldwide creates the ideal condition for virus evolution. On this respect, a new SARS-CoV-2 variant (B.1.1.7) has been recently mainly identified in the UK, and from its occurrence, it is spreading rapidly also in other countries [15]. This lineage shows an unusually large number of genetic changes, particularly in the S protein (Fig. 2). It seems that this variant acquired 17 mutations all at once (14 mutations and 3 deletions), and it has been hypothesized to have an increased likelihood to infect more rapidly and easily human subjects [15], but more data are needed to support this hypothesis. A potentially very important mutation, N501Y, detected in the S protein, seems to increase its affinity to the hACE2 receptor, and it has been already identified in over 2,057 individuals in the UK [15, 19]. Thus, it is crucial to constantly monitor all circulating forms in near real time for tracing SARS-CoV-2 epidemic and to promptly implement the proper strategies for disease containment.

This global effort was also joined by Nextstrain [20], an open-source project started in 2018 which integrates the SARS-CoV-2 genomic data updated daily (<https://github.com/nextstrain/ncov>) to interactively visualize the phylogeny of this emerging viral pathogen to better understand its dispersion dynamic and improve outbreak response [21–23]. Thanks to the genomic epidemiology and surveillance, it has been possible to trace the SARS-CoV-2 evolution worldwide. Genetic analyses and sur-

veillance allow for the characterization of circulating viral lineages, the inference of introduction events, and the reconstruction of transmission patterns. Together with epidemiological data, they constitute powerful tools to assist public health initiatives and preparedness.

Conclusion

The active monitoring of all the SARS-CoV-2 variants and/or lineages that are currently circulating and co-circulating around the world highlights the need for enhanced genomic and epidemiological surveillance worldwide in order to rapidly determine the ongoing virus evolution.

Acknowledgements

We thank all the authors and originating and submitting laboratories that have kindly deposited and shared genome data on GISAID EpiCoV database, on which this research is based.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This work was supported by CNPq (452537/2019-3). M.G. is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

References

- 1 Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: the species and its viruses: a statement of the Coronavirus Study Group. *bioRxiv*. 2020 Feb.
- 2 WHO. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020 [Internet]. [cited 2020 Dec 14]. Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>.
- 3 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 Mar;579(7798):265–9.
- 4 Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*. 2020;9(1):221–36.

- 5 Fisher D, Heymann D. Q&A: the novel coronavirus outbreak causing COVID-19. *BMC Med.* 2020 Feb;18(1):57.
- 6 World Health Organization (WHO). Coronavirus disease (COVID-19) Situation Report: 103, 2020 May 2 [Internet]. 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200502-covid-19-sitrep-103.pdf?sfvrsn=d95e76d8_4.
- 7 Horowitz J, Bubola E. On day 1 of lockdown, Italian officials Urge Citizens to Abide by rules. *NYTimes.com.* 2020 Mar.
- 8 BBC. Coronavirus: Italy imposes regional lockdown as Europe battles surges: BBC News; 2020 Nov [cited 2020 Dec 14]. Available from: <https://www.bbc.com/news/world-europe-54839429>.
- 9 BBC. Covid: how are European countries tackling the pandemic? BBC News; 2020 Dec [cited 2020 Dec 21]. Available from: <https://www.bbc.com/news/explainers-53640249>.
- 10 Financial Times. Lockdowns compared: tracking governments' coronavirus responses | Free to read [Internet]. [cited 2020 Dec 21]. Available from: <https://ig.ft.com/coronavirus-lockdowns/>.
- 11 Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell.* 2020 Aug;182(4):812–e19.
- 12 Tegally H, Wilkinson E, Lessells RR, Giandhari J, Pillay S, Msomi N, et al. Major new lineages of SARS-CoV-2 emerge and spread in South Africa during lockdown. *Infect Dis.* 2020.
- 13 Thomson EC, Rosen LE, Shepherd JG, Spreafico R, Filipe A da S, Wojcechowskyj JA, et al. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *bioRxiv.* 2020 Nov.
- 14 Hodcroft EB, Zuber M, Nadeau S, Crawford KHD, Bloom JD, Veesler D, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv.* 2020 Nov.
- 15 Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli C, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations [Internet]. *Virological.* 2020 Dec [cited 2020 Dec 20]. Available from: <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>.
- 16 GISAID Initiative [Internet]. [cited 2020 Dec 14]. Available from: <https://www.epicov.org/epi3/frontend#34f433>.
- 17 Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 2020 Nov;5(11):1403–7.
- 18 O'Toole Á, McCrone J. phylogenetic assignment of named global outbreak lineages. [Internet]. 2020. Available from: <https://github.com/hCoV-2019/pangolin>.
- 19 COG-UK. COG-UK update on SARS-CoV-2 spike mutations of special interest. 2020.
- 20 Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics.* 2018 Dec;34(23):4121–3.
- 21 Rambaut A. Phylodynamic analysis | 176 genomes | 2020 Mar 6 [Internet]. *Virological.* 2020 Jan [cited 2020 Dec 11]. Available from: <https://virological.org/t/phylogenetic-analysis-176-genomes-6-mar-2020/356>.
- 22 Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang ML, et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science.* 2020 Oct; 370(6516):571–5.
- 23 Su YC, Anderson DE, Young BE, Zhu F, Linster M, Kalimuddin S, et al. Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2. *bioRxiv.* 2020 Mar.