



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Short Communication

Characterizing SARS-CoV-2 genome diversity circulating in South American countries: Signatures of potentially emergent lineages?

Marina Muñoz^{a,*}, Luz H. Patiño^a, Nathalia Ballesteros^a, Alberto Paniz-Mondolfi^{b,c}, Juan David Ramírez^{a,*}^a Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia^b Department of Pathology, Molecular and Cell-Based Medicine, Laboratory of Microbiology, Icahn School of Medicine at Mount Sinai, New York, USA^c Instituto de Investigaciones Biomédicas IDB/Incubadora Venezolana de la Ciencia, Barquisimeto, Venezuela

ARTICLE INFO

Article history:

Received 9 January 2021

Received in revised form 15 February 2021

Accepted 16 February 2021

Keywords:

SARS-CoV-2

Lineages of epidemiological concern

B.1.1.7

B.1.351

P.1

South America

ABSTRACT

Objectives: To evaluate the genomic diversity and geographic distribution of SARS-CoV-2 lineages in South America.**Methods:** SARS-CoV-2 lineages from a public dataset of 5583 South American genome assemblies were analyzed. Polymorphisms in the main open reading frames were identified and compared to those in the main lineages of epidemiological concern: B.1.1.7 (UK) and B.1.351 (South Africa).**Results:** Across 16 South American countries, 169 lineages were identified; major lineage B had the greatest diversity and broadest geographic distribution. Seventeen predominant lineages were analyzed revealing 2 dominant lineages of concern: P.1 (Brazilian variant) and B.1.1.7 with 94 and 28 genomes, respectively, both with 33 polymorphisms (other lineages displayed ≤ 24 polymorphisms). A high number of polymorphisms were detected with a limited number of common variable positions, in common with the profile of the main lineages of epidemiological concern.**Conclusions:** The ever-increasing genetic diversity of SARS-CoV-2 continues to lead to novel lineage emergence. Various variants and lineages are now present across South America, dominated by major lineage B. The circulation of P.1 and B.1.1.7 and the high number of polymorphisms highlight the importance of genomic surveillance to determine introduction events, identify transmission chains, trace emergence, and implement prevention, vaccination and control strategies.© 2021 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Genomic surveillance, together with real-time monitoring and data-sharing networks, has become a valuable tool to improve understanding of SARS-CoV-2 transmission and epidemic dynamics in developed countries. Following the onset of the COVID-19 pandemic, multiple SARS-CoV-2 variants have arisen, including the emerging lineage B.1.1.7 initially described in the UK and now spreading globally. This lineage is of interest because of its estimated increased transmissibility (Rambaut et al., 2020a). Despite substantial advances, the implementation of genomic surveillance remains a challenge for most developing countries where access to whole genome sequencing is limited. This study aimed to characterize the genetic lineages circulating in South

America and provide a deeper understanding of the geographical pattern of distribution and potential diversification pathways of the virus across the region.

Methods

We analyzed a total of 5583 high-quality SARS-CoV-2 genomes from South America held in the publicly accessible Global Initiative on Sharing All Influenza Data (GISAID) database (Hadfield et al., 2018) as of February 8, 2021 (Supplementary Table 1). A typing report generated through Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) tool (Rambaut et al., 2020b) was downloaded and analyzed using the previously reported scheme (Ramirez et al., 2020). Results were graphically represented in Microreact (Argimon et al., 2016). A set of lineages of interest was selected for comprehensive analysis based on their predominant distribution in (first reporting date identified and frequency of the total data for that lineage) in South American countries. Genetic polymorphisms

* Corresponding authors.

E-mail addresses: claudia.munoz@urosario.edu.co (M. Muñoz), juand.ramirez@urosario.edu.co (J.D. Ramírez).

across the main Open Reading Frames (ORFs) in sequences aligned to the lineages of interest were identified and compared to the variants of concern reported worldwide (<https://cov-lineages.org/index.html>), following the previously described methodology (Ramirez et al., 2020).

Results

A total of 5583 whole-genome assemblies from 16 South American countries (Supplementary Table 1) were analyzed. The PANGOLIN tool assigned these genomes to 169 lineages, with the

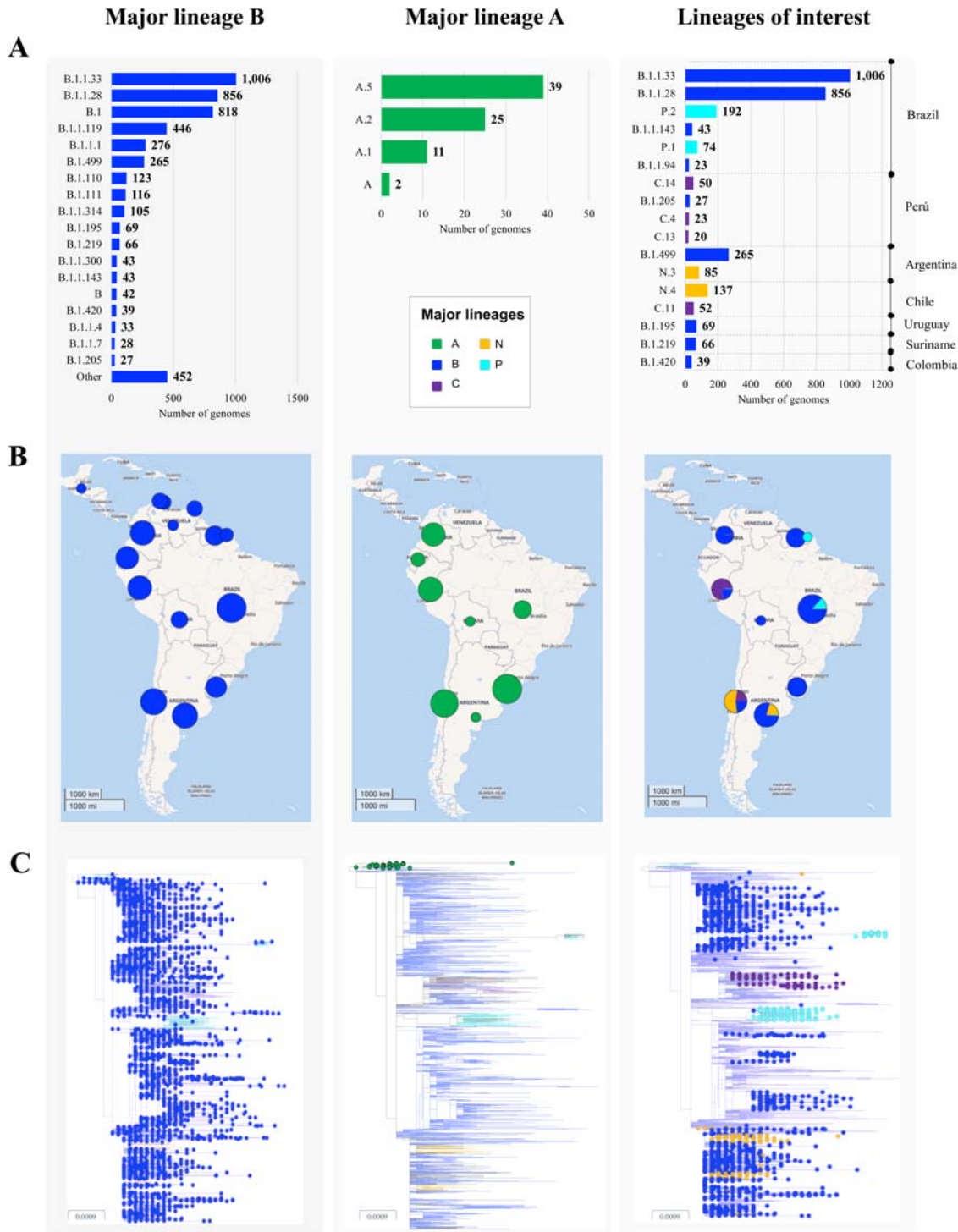


Figure 1. SARS-CoV-2 lineages reported in South American countries, defined by the PANGOLIN scheme. (A) Number of genomes reported for major lineage B (majority) and A, as well as for lineages of interest in South American countries, defined as those with predominant distribution in specific countries of the region (first reporting date identified and frequency of the total data for that lineage). For major lineage B, the lineages with <25 reports were included in the category 'Others'. (B) Profiles of the geographic distribution of the significant lineages identified in South American countries. (C) Phylogenetic relationships of the major lineages identified from the alignment of 5583 sequences with a length of 6319 positions, corresponding to the genome single-nucleotide polymorphisms without untranslated regions.

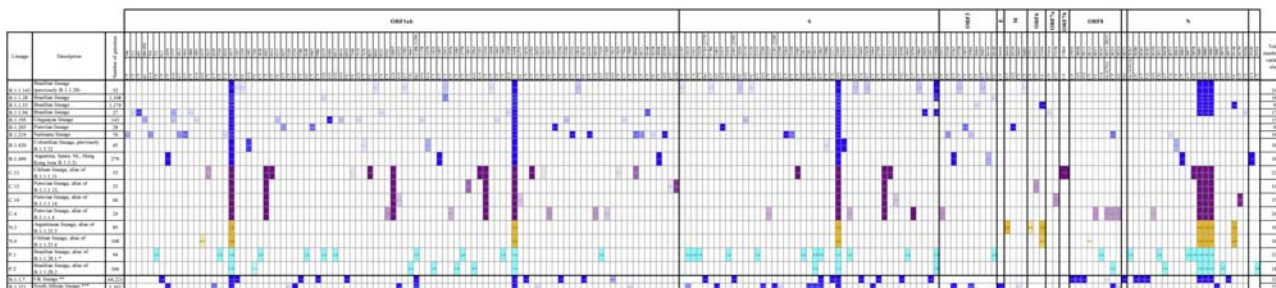


Figure 2. Variable positions in the main ORFs for SARS-CoV-2 lineages of first report in lineages of interest in South American countries compared with variants of epidemiological concern worldwide: B.1.1.7 and B.1.351. The lineages of interest were inspected (upper part). The variable sites were compared with those found in the variants of concern report B.1.1.7 and B.1.351 (bottom). Colors were assigned by major lineage according to the color code of Figure 1. The number inside the boxes indicates the proportion of genomes exhibiting polymorphism compared to the reference strain NC045512-2-Wuhan-Hu-1. The absence of color represents a conserved position. The terms “Del” and “Ins” indicate a deletion and insertion, respectively. The description of lineages was consulted in <https://cov-lineages.org/lineages.html>.

*Lineage with a number of spike mutations with likely functional significance E484K, K417T, and N501Y; described in <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>.

**Lineage of epidemiological concern associated with the N501Y mutation. More information can be found at cov-lineages.org/global_report.html.

***Defined as the new variant 501Y.V2-The description of this lineage is available as a preprint: <https://www.medrxiv.org/content/10.1101/2020.12.21.20248640v1>.

majority (150) under major lineage B and the others under 5 significant lineages: A (4 lineages, 77 genomes), C (8 lineages, 152 genomes), I (1 lineage, 3 genomes), N (4 lineages, 232 genomes) and P (2 lineages, 266 genomes). A set of 17 lineages of interest were comprehensively analyzed based on their predominant distribution in South America: 6 in Brazil, 4 in Peru, 2 in Chile and Argentina, and 1 in Uruguay, Suriname and Colombia (Figure 1A). The geographical distribution of these lineages is represented in Figure 1B.

Lineage B.1.1.7 (UK variant) was significant, with 28 reports from the following countries: Brazil (n = 19), Ecuador (n = 6) and Argentina, Peru, and Trinidad and Tobago (1 report per country). Lineage P.1 (Brazilian variant) was also identified, with 74 genomes. The South African lineage B.1.351 has not been reported in the region at the analysis date. The phylogenetic reconstructions confirm the high diversity of the major lineage B and its close relationship with all other major lineages in monophyletic clusters (Figure 1C). The B lineages are predominant in some South American countries; however, the global analysis of their distribution indicates that they have been reported in other countries only during the advance of the pandemic (Supplementary Table 2). Lineage B.1.1.5 was previously the most frequently detected lineage in Colombia (13/88 = 14.8%) but had a limited frequency elsewhere in the world (44/2744 = 1.6%) (Ramirez et al., 2021a), it is now called B.1.420, and there are reports of circulation in other countries (Supplementary Table 2).

In the lineages of interest, 194 polymorphic sites were detected in the main ORFs (Figure 2); 3 were shared by all the genomes evaluated: 2 in ORF1ab at positions 3037 (C/T) and 14 408 (C/T), and 1 in the S gene at position 23 403 (A/G). A set of 3 consecutive positions in the N gene (28,881–28,883) was found in most lineages, absent only in 50% of the B lineages. Some variable positions were found in common to major lineages, as is the case of positions 4002, 10 323 and 13 536, in the ORF1ab, as well as 23 731 in the S gene, which was found only in the major lineage C. The lineages included in this group are predominant in Chile (n = 1) and Peru (n = 3). Variable sites common to the 2 major N lineages were found at position 27 299 of ORF6 (T/C) and 29 148 of N gene (T/C). Other variable positions were unique to each lineage evaluated and, in most cases, were present across all the genomes evaluated in that lineage. The lineages of reported concern had the highest number of polymorphisms compared with other evaluated lineages, with 33 for both P.1

(Brazilian variant) and B.1.1.7 (UK variant), and 27 for B.1.351 (South African variant).

Discussion

Identifying the diversity of SARS-CoV-2 is essential for monitoring the dynamics of the pandemic’s dispersion in different regions of the world. A high diversity of lineages was found in the South American region, and some lineages with a predominant distribution in particular countries have now spread to other countries in the region.

Of particular note are the rapid increase in the number of sequences of the P.1 variant (a variant with suspected increased transmissibility that was initially detected and described in Manaus, Brazil) and the occurrence in the lineages of interest of polymorphisms across the main ORFs of the genome in common with polymorphisms in the B.1.1.7 variant of concern (identified in Brazil, Ecuador, Argentina, Peru, and Trinidad and Tobago). These findings provide insight into the potential emerging lineages across South America resulting from the changing nature of its genome, leading to the appearance of new variants, a frequent event in this type of virus (Lauring and Hodcroft, 2021).

The potential impact of these genomic variants on infectivity and pathogenicity remains to be fully determined. However, it has been proposed that some of the polymorphisms found may have potential functional significance, mainly those located in the spike gene (i.e. N501Y, E69/70, P681H, 144Y, A570D, E484K, K417N, K417T), which could impact transmission, and raises concern about immunological escape (Rambaut et al., 2020a). The broad SARS-CoV-2 lineage diversity circulating in South America could exacerbate the impact of the pandemic in the region, affecting the reliability of molecular diagnosis schemes (Ramirez et al., 2021b) and even the efficacy of vaccines since the polymorphisms in the spike gene could promote immune evasion (Bouayad, 2020, McCarthy et al., 2021).

The phylogenetic diversity of SARS-CoV-2 circulating in South America represents a dynamic and fast-growing number of lineages in the region. However, the limited number of available genomes for the region compared with others prevents a clear view of the overall genetic landscape, reaffirming the need to strengthen genomic-based surveillance systems. In future, it is necessary to develop studies to evaluate the impact of this lineage diversity on the reliability of molecular diagnosis and the success of vaccination strategies.

Conflict of interest

None declare.

Funding source

This work was funded by DIRECCIÓN DE INVESTIGACIÓN E INNOVACIÓN from Universidad del Rosario.

Ethical approval

Not applicable as all the information was downloaded from GISAID.

Acknowledgements

We thank the High Computing Cluster (CENTAURO) from Universidad del Rosario for their support during the data analysis.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijid.2021.02.073>.

References

- Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016;2(11):e000093.
- Bouayad A. Innate immune evasion by SARS-CoV-2: comparison with SARS-CoV. *Rev Med Virol* 2020;30(6):1–9.
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34(23):4121–3.
- Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2—what do they mean?. *JAMA* 2021;325(6):529–31.
- McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 2021;71(6534):1139–42.
- Rambaut A, Holmes EC, O’Toole A, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020a;5(11):1403–7.
- Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *nCoV-2019 Genomic Epidemiol* 2020b; Preprint.
- Ramirez JD, Munoz M, Hernandez C, Florez C, Gomez S, Rico A, et al. Genetic diversity among SARS-CoV2 strains in South America may impact performance of molecular detection. *Pathogens* 2020;9(7).
- Ramirez JD, Florez C, Munoz M, Hernandez C, Castillo A, Gomez S, et al. The arrival and spread of SARS-CoV-2 in Colombia. *J Med Virol* 2021a;93(2):1158–63.
- Ramirez JD, Munoz M, Patiño LH, Ballesteros N, Paniz-Mondolfi A. Will the emergent SARS-CoV2 B.1.1.7 lineage affect molecular diagnosis of COVID-19?. *J Med Virol* 2021b;. doi:<http://dx.doi.org/10.1002/jmv.26823> Online ahead of print.