



The arrival and spread of SARS-CoV-2 in Colombia

Juan David Ramírez¹  | Carolina Florez² | Marina Muñoz¹ | Carolina Hernández¹ | Adriana Castillo¹ | Sergio Gomez² | Angelica Rico² | Liseth Pardo² | Esther C. Barros² | Sergio Castañeda¹ | Nathalia Ballesteros¹ | David Martínez¹ | Laura Vega¹ | Jesús E. Jaimes¹ | Lissa Cruz-Saavedra¹ | Giovanni Herrera¹ | Luz H. Patiño¹ | Aníbal A. Teherán³ | Ana S. Gonzalez-Reiche⁴ | Matthew M. Hernandez⁴ | Emilia Mia Sordillo⁴ | Viviana Simon⁴ | Harm van Bakel⁴ | Alberto Paniz-Mondolfi⁴ 

¹Grupo de Investigaciones Microbiológicas-UR (GIMUR), Departamento de Biología, Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia

²Instituto Nacional de Salud, Bogotá, Colombia

³Fundación Juan N Corpas, Bogotá, Colombia

⁴Icahn School of Medicine at Mount Sinai, New York, New York

Correspondence

Juan David Ramírez, Grupo de Investigaciones Microbiológicas-UR (GIMUR), Departamento de Biología, Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia.
Email: juand.ramirez@urosario.edu.co

Funding information

Universidad del Rosario,
Grant/Award Number: Internal funds

Abstract

We performed phylogenomic analysis of severe acute respiratory syndrome coronavirus-2 from 88 infected individuals across different regions of Colombia. Eleven different lineages were detected, suggesting multiple introduction events. Pangolin lineages B.1 and B.1.5 were the most frequent, with B.1 being associated with prior travel to high-risk areas.

KEYWORDS

B.1, Colombia, COVID-19, lineage, SARS-CoV-2

1 | INTRODUCTION

The number of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections is rapidly increasing throughout South America. The first coronavirus disease 2019 (COVID-19) case in South America was reported on 25 February 2020, in Sao Paulo, Brazil, and since has spread so that as of 20 March, cases of SARS-CoV-2 infection have been reported from all Latin American countries. The healthcare systems in the region are fragile, including multiple social and economic issues that together might be devastating for this vulnerable region.¹ Therefore, genomic epidemiology studies must be in place to understand the dynamics of COVID-19 epidemic.

Colombia, the fifth largest country in South America is the fourth country in number of confirmed new COVID-19 cases in the region by 30th July 2020. Following identification of the first COVID-19 case in Colombia, on 6 March 2020, in a female traveler returning from Milan, Italy, the Colombian government implemented early control measures. An ongoing mandatory lockdown and travel ban/

restriction was put in place on 23 March 2020, which included the closure of all the airports across the country. Although these containment measures certainly helped reduce the basic reproduction number (R_0) from 4.8 to 2.2, they have been unable to fully limit SARS-CoV-2 spread in Colombia.²

Colombia's geographic location makes it an important crossroad in the Andean region, attracting many travelers across South America and making it a particular vulnerable region with important implications for internal spread and dissemination to its multiple bordering countries. However, in-depth studies on the molecular epidemiology and SARS-CoV-2 strains circulating in Colombia and elsewhere in South America are still lacking.

2 | METHODS AND RESULTS

Individuals meeting case-definition criteria established by the Colombian Ministry of Health and Social Protection were screened for

SARS-CoV-2 infection at different hospitals and healthcare centers in 16 of the 32 Departments of Colombia between 31st March and 1st May 2020.³ Molecular detection of SARS-CoV-2 in clinical nasopharyngeal swab (NP-VTM) specimens was performed using the Berlin Charité protocol,⁴ with 88 positive cases that were randomly selected for further study. Most of the SARS-CoV-2 infected patients were identified in four (Andean, Caribbean, Pacific, and Orinoco) of the six geographic regions of Colombia; particularly in the Departments of Valle del Cauca (35.9%), Cundinamarca (11.2%), Boyacá (10.1%), Antioquia (8.9%), and Huila (7.8%). Sociodemographic characteristics of the 88 SARS-CoV-2-positive patients showed that the average age was 44 (ranging from 36-58), with 58% ($n = 51$) male and 42% ($n = 37$) female subjects. Different risk factors for exposure were identified, 12 (13.6%) patients were health care workers, 55 (62%) had close contact with infected patients and 23 (28.4%) had traveled to high-risk areas (Mostly European countries). On presentation 17 (19.3%) were asymptomatic, 71 (80.7%) were symptomatic, and 26 (29.5%) required hospitalization. At presentation, the most common symptoms were respiratory (80.6%), fever (59%), and gastrointestinal symptoms (33%). Respiratory symptoms ranged from nonspecific influenza-like symptoms (dry cough and shortness of breath) to respiratory failure (5.7%). Twenty-nine (33%) patients had concurrent conditions, such as diabetes, hypertension, COPD, asthma, cardiac failure, and cancer.

To assess the genetic diversity and origins of SARS-CoV-2 in Colombia, we sequenced and assembled viral genomes from total RNA extracted from NP-VTM clinical specimens. Sample preparation for sequencing was done using whole-genome amplification with custom designed tiling primers and the Artic Consortium protocol (<https://artic.network/ncov-2019>), with modifications as reported elsewhere using an Illumina MiSeq instrument.⁵ Comparative genome analysis of our 88 cases and three previously reported Colombian cases was carried out relative to publicly available background data from 2744 cases sampled from the GISAID EpiCoV database to obtain a full representation of global lineage diversity.⁶ Lineage assignments were performed using the Phylogenetic Assignment of Named Global Outbreak Lineages tool "Pangolin." Consensus viral sequences from each case were also submitted to GISAID (accessions: EPI_ISL_447734-EPI_ISL_447817).

Whole genome sequences for the samples included in each dataset were aligned using MAFFT v7.40755⁷ with FFT-NS-2 algorithm and default parameter settings. All multiple sequence alignments were manually curated to remove 5'- and 3'-untranslated regions as potentially ambiguous regions, and then analyzed in trimAl to remove spurious or poorly aligned sequences. The best substitution model was chosen for all alignments in jModelTest2 v0.1.1 58. Maximum likelihood (ML) trees were inferred using IQtree2 v1.6.1 using the best substitution model, default heuristic search options, and ultrafast bootstrapping with 1000 replicates and other parameters by default. The ML reconstruction showed that the Colombian genomes have a close phylogenetic relation to a wide range of SARS-CoV-2 strains across 11 different Pangolin lineages⁷, with a predominance of B lineages all across the country (Figure 1). We performed

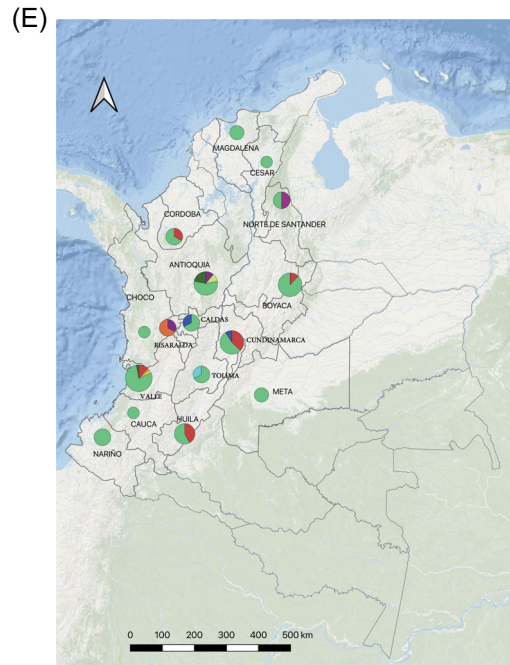
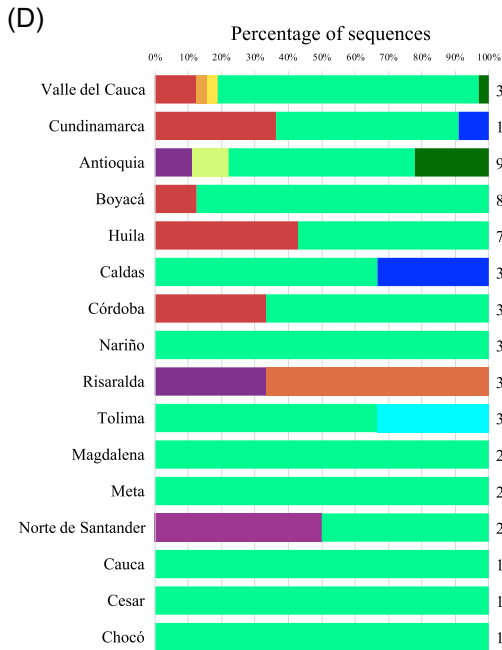
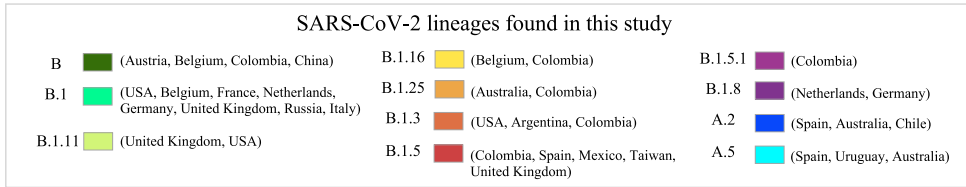
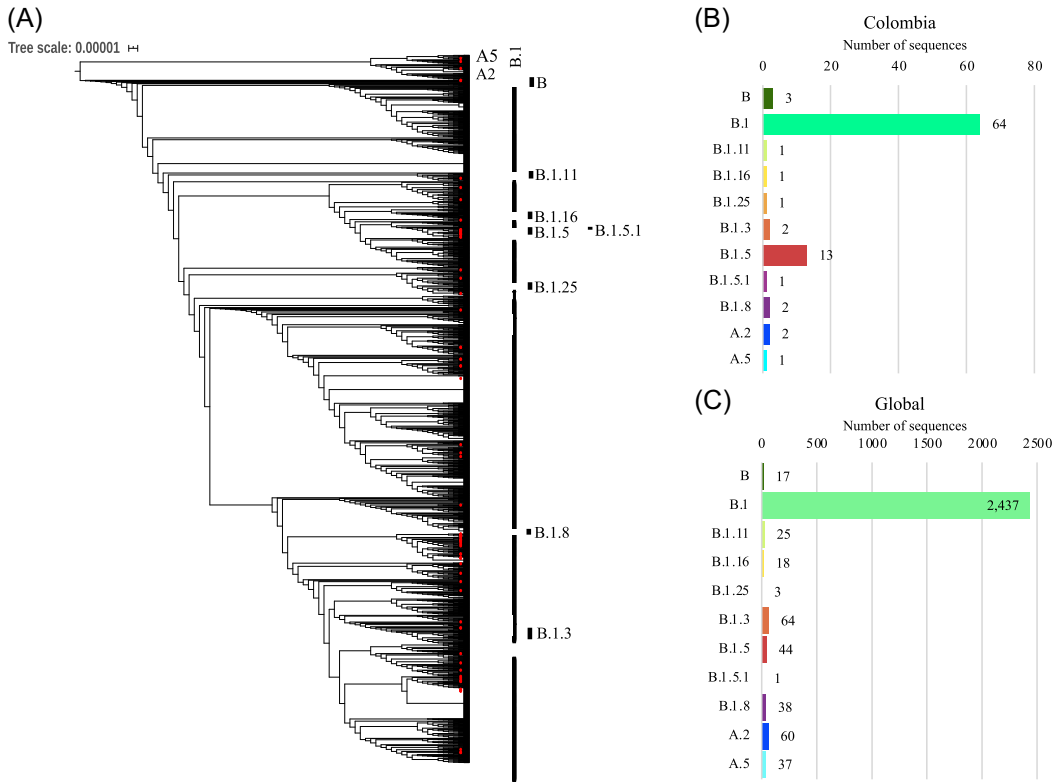
univariate analysis to determine whether certain lineages were associated with a health-care worker status, hospital exposure (including intensive care unit admission), or a travel history to high-risk areas. A significant association ($P = .033$) was only found between infection with B.1 lineage and a travel history to high-risk areas. In conclusion, Pangolin lineage B.1 was associated with prior travel to high-risk areas.

We further constructed a time-scaled ML phylogeny using TreeTime and specimen collection date constraints.⁸ At least nine potential introductions during the dispersion of SARS-CoV-2 into the country were identified between 20th January (confidence interval [CI], 95% 18th-20th January) and 12th March (CI, 95% 11th-12th March) (Figure 2). Finally, we evaluated the substitution in the Spike protein, the single preliminary associated with a phenotypical effect in the SARS-CoV-2 genome.⁹ This was conducted by inspection of the whole genome alignment of the analyzed genomes (88 Colombian genomes and the Wuhan reference strain (NC_045512.2) at position 23 403, using the Unipro UGENE v.35 bioinformatics toolkit.¹⁰ The G-to-A mutation in Spike protein (at position 23 403 in the Wuhan reference strain NC_045512.2) was found in 85 of the 88 Colombian genomes sequenced in this study. This substitution was recently recognized as biologically significant and associated with more transmissible populations of SARS-CoV-2.⁹

3 | DISCUSSION

SARS-CoV-2 has become one of the most important epidemics in the last half-century with over four million infected individuals and close to 300 000 reported deaths worldwide. To date, most affected countries are industrialized nations with robust public health systems and state-of-the-art medical facilities, which despite this have undergone severe strains in the course of the pandemic. On the other hand, as the spread wave attenuates in many regions of the world, other regions, such as South America are still witnessing an important rise in the number of cases. Many developing countries have been particularly affected by this novel coronavirus. Brazil, Ecuador, Peru, and Colombia are clear examples where COVID-19 is taking a deadly toll.

These countries share in common several factors, such as marked poverty, lack of access to water, sanitation, and adequate medical facilities as well as distrust in public governance.¹¹ These aspects are relevant in understanding the course of the epidemics and how they can affect transmission dynamics of the virus. For example, the crowded nature of slums and suburban settings in larger cities of South America preclude adequate social distancing efforts. In addition, the unfeasibility of complying with quarantine measures due to concern over loss of income or losing employment, as well as the lack of water and inappropriate wastewater treatment may favor transmission and halt potential mitigation efforts needed for containment of the virus. This is why alongside strengthening public health capabilities there is an urgent need to better understand and address the main drivers influencing the epidemic spread in this



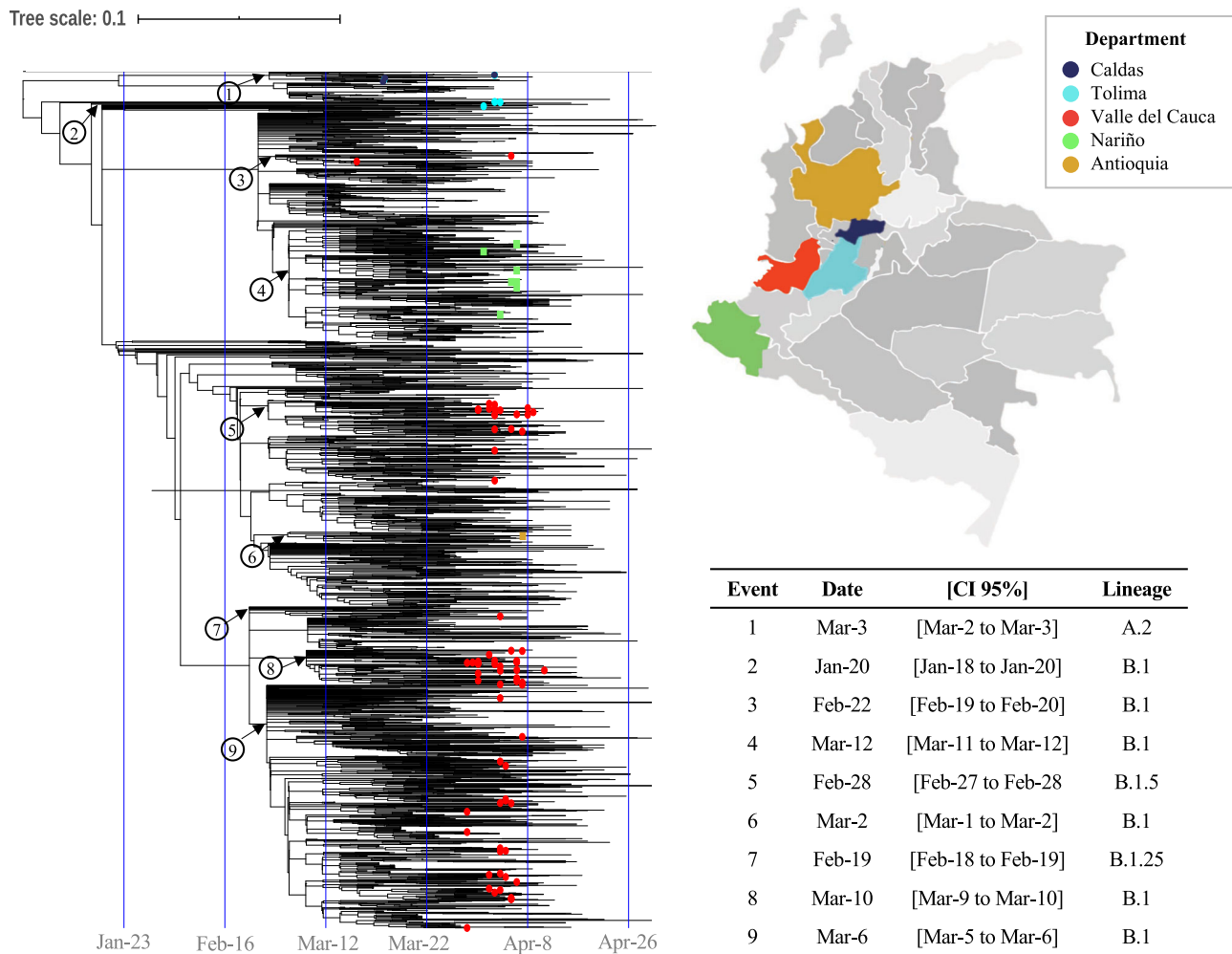


FIGURE 2 Multiple and early introductions of SARS-CoV-2 lineages in Colombia. Time-scaled tree built in TreeTime from the trimmed whole genome alignment of the global background diversity (left). The colored dots indicate the 91 genomes encompassing the 11 lineages herein described. Dots are colored according to the introductions to specific geographical regions of Colombia shown on the right and labeled with numbers in the tree (Turquoise = Tolima, Red = Valle del Cauca, Yellow = Antioquia, Purple = Caldas, and Green = Nariño). The nodes with the dates estimates are indicated with the blue arrows and the number of introductions with the numbers and time estimates in the table (right). The displayed time tree was inferred under a strict clock model with a fixed substitution rate of 0.8×10^{-3} , based on previous rate value estimates.⁵ TreeTime analyses were run for a total of six iterations and marginal date estimates of ancestral states are shown with 90% confidence intervals. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

complex scenario. Assessing these scenarios with cutting-edge technologies that allow tracking this novel pathogen in high-risk geographical regions, make genome sequencing and phylogenetics fast and reliable analytic tool for addressing the main epidemiological trends and improve outbreak response and contention.

Our study represents the first overview on the molecular epidemiology of SARS-CoV-2 in Colombia. The genome sequence of 88 samples isolated from patients with COVID-19 from different regions of Colombia allowed identifying 11 independent lineages, suggesting a massive introduction of the virus in the country (Figure 1).

FIGURE 1 Distribution of several severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lineages in Colombia. A, Maximum likelihood (ML) tree of 2744 global background and 91 Colombian genomes (red dots; including three genomes previously deposited in GISAID), the tree was rooted with Pangolin coronavirus (MT084071.1). The pangolin nomenclature is used to show the eleven lineages detected in the country. B, Percentage of Colombian genomes assigned to a specific Pangolin lineage. C, Percentage of genomes from the global diversity (highlighting their geographical origin) assigned to the lineages described in Colombia using Pangolin nomenclature D, Percentage of genomes belonging to the different lineages by Department of Colombia, assumed as geographical regions according with the national administrative and political division. E, The geographical distribution of SARS-CoV-2 lineages in Colombia using IQGIS

The most widespread lineage found across all departments was B.1, which has been reported in at least 43 different countries following a cosmopolitan distribution.⁶ Our data supports variable sources of introductions of the SARS-CoV-2 into Colombia predominantly from Europe and North America (Figure 1), as well as active transmission despite establishment of early containment measures before the date of the first case detected and evidenced as multiple entries into multiple regions (Figure 2). Three scenarios may explain such trend. First, infected travelers and migrants from countries already affected by the SARS-CoV-2 likely entered Colombia before the travel ban and closure of Colombia's borders (23rd March 2020). Second, many Colombian citizens with limited economic and/or social resources have been unable to comply with quarantine measures. Third, vast variations in ethnicity, climate, and sociodemographic features (environmental heterogeneity) across Colombia may be influencing the presentation and spread of the virus.

Despite the majority of identified lineages are European and North American in origin, there were a few of them from other different geographical locations. Our data suggests (based on the initial description of these lineages) additional introductions from China, Australia, United States, Canada, Chile, and Iceland lineages (Figure 1). That is the case of lineages A.5, B.1.3, B.1.11, B.1.5.1, and B.1.25. Interestingly, the third most common lineage in the country was the B lineage, which has been reported worldwide in returning travelers from China, and which may suggest an independent introduction in the country of the ancestral SARS-CoV-2 lineage. This is also the case for other Latin-American countries, such as Uruguay, which has reported the occurrence of B, B.1 and A.1a lineages,¹² Chile that has reported B and A.2a lineages¹³ and Brazil with the largest genomic epidemiology study so far in the region.¹⁴ Despite installment of early contention measures like closing borders and international airports on 23rd March (17 days after the first case was detected), our data suggests such actions were not enough to avoid multiple introductions of the virus into Colombia. Consider in addition that El Dorado international airport (the biggest airport in the country) and the second most transited in Latin America could have influenced the early spread of SARS-CoV-2 in the country but also across the region. Particularly interesting, we observed some departments, such as Valle del Cauca, Cundinamarca, and Antioquia (Figure 1) that had circulation of different lineage that is supported by the fact that these departments have the most populated cities in the country.

The arrival of SARS-CoV-2 in South America poses particular challenges, as the virus now spreads across a region with diverse and complex geopolitical and sociocultural contexts. Marked poverty, urban and suburban overcrowding, scarce sanitary conditions as well as overwhelmed public health systems sharply contrast with the way SARS-CoV-2 has impacted most of the industrialized countries of the world so far. Such conditions may negatively impact viral dynamics favoring transmission and long-term persistence. In fact, the World Health Organization has recently stated that South America has now become the new epicenter of the global coronavirus pandemic, thus urging implementation of widespread population surveillance (including genomic epidemiology studies) and reinforcing containment measures.

It is well known that mutational events in the S and N genes of coronaviruses may affect its pathogenicity.¹⁵ In fact, it has been demonstrated that both of these genes are undergoing episodic selection as the virus is transmitted amongst humans.¹⁶ More recently, the description of several mutations in the spike glycoprotein of SARS-CoV-2 capable of inducing a missense mutation has suggested the possibility for increased viral infectivity and virulence.^{9,17} Also, illustrating how the two S1 domains recognize different receptors and how the spike proteins are regulated to undergo conformational transitions and increase infectivity in coronaviruses. To date there are few studies that have associated the SARS-CoV-2 lineage with the infection severity¹⁸ and differential diagnosis.¹⁹ In our case, we identified the key mutation D614G in most of the Colombian genomes which has been associated with increased infectivity,⁹ future studies should unveil its clinical consequences and impact in South America and its regional particularities.

In conclusion, this represents the first genomic epidemiology study of SARS-CoV-2 in Colombia. Future studies in the country and elsewhere in South America, including sequencing of viral genomes as the predicted epidemic peak approaches, and of contact cases and spread clusters, may help to better identify transmission routes and inform potential prevention measures. Our study supports the relevance of genomic surveillance and the critical need to establish coordinated efforts to generate genomic data in South America that will enable integrative analyses to uncover SARS-CoV-2 dynamics at the continental level.

ACKNOWLEDGMENTS

The authors thank Dirección de Investigación e Innovación from Universidad del Rosario for funding this study. Funding was provided by the University of Glasgow, Scottish Funding Council and the Global Challenges Research Fund (GCRF) and GCRF Research Network EP/T003782/1.

AUTHOR CONTRIBUTIONS

JDR, CF, and APM designed the study and wrote the manuscript. MM, CH, AC, SC, NB, DM, LV, JEJ, LS, and GH. AT collected and analyzed the data. ASG, MMH, EMS, VS, and HB conducted the sequencing. All authors approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data are available in GISAID nextstrain.

ORCID

Juan David Ramírez  <http://orcid.org/0000-0002-1344-9312>

Alberto Paniz-Mondolfi  <http://orcid.org/0000-0003-1259-1736>

REFERENCES

- Rodriguez-Morales AJ, Gallego V, Escalera-Antezana JP, et al. COVID-19 in Latin America: the implications of the first confirmed case in Brazil. *Travel Med Infect Dis.* 2020;35:101613.
- Datos de Coronavirus Colombia. Ministerio de salud y protección social. 2020. https://www.minsalud.gov.co/salud/publica/PET/Paginas/Covid-19_copia.aspx
- Saavedra Trujillo CH. Resumen: consenso colombiano de atención, diagnóstico y manejo de la infección por SARS-COV-2/COVID-19 en

- establecimientos de atención de la salud—recomendaciones basadas en consenso de expertos e informadas en la evidencia. *Infectio*. 2020;24:3.
4. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR 2020. *Eur. Surveill*. 2020;25: 2000045.
 5. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science*. 2020.
 6. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *Nature Microbiol*. 2020.
 7. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016;2(1):07.
 8. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform*. 2016;20:4.
 9. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. 2020.
 10. Okonechnikov K, Golosova O, Fursov M. the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28: 1166-1167.
 11. Miller MJ, Loaiza JR, Takyar A, Gilman RH. COVID-19 in Latin America: novel transmission dynamics for a global pandemic. *PLOS Negl Trop Dis*. 2020;14(5):e0008265.
 12. Salazar E, Kuchipudi SV, Christensen PA, et al. Multiple introductions, regional spread and local differentiation during the first week of COVID-19 epidemic in Montevideo, Uruguay. *BioRxiv*. 2020. <https://doi.org/10.1101/2020.05.09.086223>
 13. Rodriguez-Morales AJ, Rodriguez-Morales AG, Méndez C, Hernández-Botero S. Tracing new clinical manifestations in patients with COVID-19 in Chile and its potential relationship with the SARS-CoV-2 divergence. *Curr Trop Med Rep*. 2020: 1-4.
 14. Candido DS, Claro IM, de Jesus JG, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. 2020.
 15. Baric RS, Yount B, Hensley L, Peel S, Chen W. Episodic evolution mediates interspecies transfer of a murine coronavirus. *J Virol*. 1997; 71(3):1946-1955.
 16. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. *BioRxiv*. 2020;92:455-459.
 17. Yao C, Bora SA, Parimon T, et al. Patient-derived mutations impact pathogenicity of SARS-CoV-2. *MedRxiv*. 2020.
 18. Chu H, Chan JFW, Yuen TTT, et al. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *Lancet Microbe*. 2020;1:e14-e23.
 19. Zhang X, Tan Y, Ling Y, et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature*. 2020;583:437-440.

How to cite this article: Ramírez JD, Florez C, Muñoz M, et al. The arrival and spread of SARS-CoV-2 in Colombia. *J Med Virol*. 2021;93:1158-1163. <https://doi.org/10.1002/jmv.26393>