Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2

Katherine Laiton-Donato[a,*], Carlos Franco-Muñoz[a*], Diego A. Álvarez-Díaz[a], Hector Alejandro Ruiz-Moreno[a], Jose A. Usme-Ciro[a,b], Diego Andrés Prada[a], Jhonnatan Reales-González[a] Sheryll Corchuelo[a], María T. Herrera-Sepúlveda[a], Julian Naizaque[a], Gerardo Santamaría[a], Jorge Rivera[a], Paola Rojas[a], Juan Hernández Ortiz[c], Andrés Cardona[c], Diana Malo[d], Franklin Prieto-Alvarado, Fernando Ruiz Gómez[e], Magdalena Wiesner[a], Martha Lucia Ospina Martínez[f], Marcela Mercado-Reyes[a].

[a] Group Genomics of Emerging Microorganisms. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá, Colombia.

[b] Centro de Investigación en Salud para el Trópico - CIST, Facultad de Medicina, Universidad Cooperativa de Colombia, Santa Marta, Colombia.

[c] Laboratorio Genómico One Health. Medellín, Colombia.

[d] Dirección de Vigilancia en Salud Pública, Instituto Nacional de Salud, Bogotá, Colombia.

[e] Ministerio de Salud y Protección Social de Colombia. Bogotá Colombia.

[f] Dirección General, Instituto Nacional de Salud, Bogotá, Colombia.

[*] These authors contributed equally.

Corresponding author: Katherine Laiton-Donato. Group Genomics of Emerging Microorganisms, Department of Public Health Research, Instituto Nacional de Salud, Av. Calle 26 No 51-20, Bogotá 111321, Colombia Tel.: (+571) 3142905064. Email: kdlaitond@unal.edu.co

Abstract

SARS-CoV-2 genetic diversity has the potential to impact the virus transmissibility and the escape from natural infection- or vaccine-elicited neutralizing antibodies. Here, we report the emergence of the B.1.621 lineage, considered a variant of interest (VOI) with the accumulation of several mutations affecting the Spike protein, including the amino acid changes I95I, Y144T, Y145S and the insertion 146N in the N-terminal domain, R346K, E484K and N501Y in the Receptor Binding Domain and P681H in the S1/S2 cleavage site of the Spike protein. The rapid increase in frequency and fixation in a relatively short time in some cities that were near the theoretical herd immunity suggests an epidemiologic impact. Further studies will be required to assess the biological and epidemiologic roles of the substitution pattern found in the B.1.621 lineage.

Keywords: SARS-CoV-2, variant of interest (VOI), evolution, emergence

Introduction

In September 2020, SARS-CoV-2 variants of concern (VOC) and variants of interest (VOI) started to be reported, with more distinctive mutations than expected from the characteristic clock-like molecular evolution of this virus evidenced during the first year pandemic (1,2). Despite mutations spanning the whole genome, an interesting feature of these emerging variants has been the presence of several amino acid substitutions falling in the Spike protein, the viral protein responsible for receptor binding and membrane fusion and also the main target for neutralizing antibodies (3). Monitoring the emergence of new variants of SARS-CoV-2 is a priority worldwide, as the presence of certain non-synonymous substitutions and INDELs could be related to biological properties, such as altering the ligand-receptor affinity, the efficiency of neutralization by naturally acquired polyclonal immunity or post-vaccination antibodies and transmission capacity (4–6).

In Colombia, the National Genomic Characterization Program led by the Instituto Nacional de Salud has carried out real-time monitoring of the SARS-CoV-2 lineages since the beginning of the pandemic (7,8). Until December 2020, over thirty lineages were circulating inside the country without evidence of VOC and VOI importation. However, a lineage turnover accompanied the third epidemic peak during March and April 2021, involving the emergence of B.1 lineage descendants with high mutation accumulation (B.1.621 and the provisionally assigned B.1+L249S+E484K) (9), as well as the introduction of the B.1.1.7, P.1 and VOI in some cities.

In this study, we reported the emergence and spread of the novel B.1.621 lineage of SARS-CoV-2, a new VOI with the insertion 146N and several amino acid substitutions in the Spike protein (Y144T, Y145S, R346K, E484K, N501Y and P681H).

Methods

Samples were selected from routine surveillance in all departments based on representativeness and virologic criteria (10). Respiratory samples were used for automated RNA extraction with magnetic beads and the RNA extracts were processed by using the amplicon sequencing protocol v3 (https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye). The

assembly of raw NGS data was performed by following the pipeline described for Oxford Nanopore Technologies (ONT) platform (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html).

Lineage assignment started by filing a new issue in the pango-designation repository (https://github.com/cov-lineages/pango-designation/issues/57) followed by designation as B.1.621 lineage by the Pangolin curation team and PangoLEARN model training for subsequent automatic lineage assignment.

Maximum likelihood tree reconstruction was performed with the GTR+F+I+G4 nucleotide substitution model using IQTREE. Branch support was estimated with an SH-like approximate likelihood ratio test (SH-aLRT). Recombination detection was performed using RDP4 software with RDP, GENECONV, Bottscan, Maxchi, Chimaera, SiSscan, and 3Seq tests (P-value < 0.05). Dataset 1 included Colombian SARS-CoV-2 sequences representative of the different lineages and dataset 2 included sequences previously reported as VOC or VOI. Adaptive evolution analysis at the codon level was estimated by Hyphy using stochastic evolutionary models. The detection of individual sites was performed with methods such as MEME (Mixed Effects Model of Evolution), and FEL (Fixed Effects Likelihood) (P-value <0.3).

Results

Lineage B.1.621 has become predominant in some departments of Colombia

The routine genomic surveillance of SARS-CoV-2 in Colombia was reinforced in January 2021 for a higher sensitivity monitoring of the potential importation of VOC. By May 7, 2021, a total of 908 sequences from Colombia were available in the GISAID database. Lineage B.1 is the best-represented lineage (with 229 records) due to its higher frequency from the beginning of the pandemic. The recently designated B.1.621 lineage has been increasingly detected from January 11, 2021 (collection date of the first genome belonging to the lineage) to date (77 records), occupying the fifth place in frequency(Figure 1A), rapidly becoming fixed in some departments located in the North of the country or co-circulating with others lineages in Bogotá D.C. (Figure 1B). The genetic background of the B.1.621 lineage includes some convergent amino acid changes previously found in several VOI and VOC.

The original assignment through the Pangolin algorithm for this monophyletic group was the B.1 lineage. However, a large number of distinctive synonymous and non-synonymous substitutions, including the following amino acid changes I95I, Y144T, Y145S in the N-terminal domain; R346K, E484K and N501Y in the Receptor Binding Domain and P681H in the S1/S2 cleavage site (Table 1) and the insertion 146N (supplementary table 1) in the Spike protein.

Lineage B.1.621 emerged from the parental B.1 lineage circulating in Colombia

The close phylogenetic relationship of the SARS-CoV-2 sequences belonging to the B.1.621 lineage with other sequences from representative lineages circulating worldwide and those circulating in Colombia suggested a recent origin from the parental lineage B.1 (Figure 2a), which was corroborated through the lineage designation (https://github.com/cov-lineages/pango-designation/issues/57) (supplementary table 2). B.1.621 lineage has recently spread to fourteen departments, with a major representation in the Caribbean region of Colombia (Figure 2b) (https://microreact.org/project/5CAiK3qCMaEgE4vYkKVpZW/b7113efc). No recombination events were found throughout the whole genome (data not shown). At least 9 codons in the Spike protein displayed a signal suggestive of positive selection (supplementary table 3).

Discussion

All genomes belonging to the B.1.621 lineage available until the end of April 2021 were included in this study, with the earliest collection date on January 11th. 2021, corresponding to a sample collected in the department of Magdalena, Colombia (EPI_ISL_1220045). While a very high and unexplained genetic distance is found between every B.1.621 sequence and all closely related sequences of the parental B.1 lineage, the whole branching pattern and intra-lineage distance suggest low diversification that can be explained by its recent origin. The spread in the country early during the third peak of the pandemic could be explained by a combination of factors, including social exhaustion as well as the genetic background of the emerging lineage, leading to changes in transmission.

In Colombia, the current strategy for SARS-CoV-2 genomic surveillance includes sampling in principal and border cities, in special groups of interest, in patients with distinctive clinical features and severity, and finally in community transmission with an unusual increase in cases.

The high frequency of the emerging B.1.621 lineage also could be related to the strengthening of the SARS-COV-2 genomic surveillance during the third peak of the pandemic in Colombia, nevertheless, it is expected to characterize an approximate 1% of the cases and determine the adjusted frequency of the lineage in the country and to evaluate the possible predominance and the replacement of other lineages in the country. For this, intensified genomic characterization will be carried out with a multi-stage sample design throughout the national territory.

During the first quarter of 2021, several convergent substitutions have been evidenced in the different lineage of SARS-CoV-2 explained by a high rate of the naive population that has explored wide options of genetic variability, the other scenario is explained by the selection pressure by monoclonal antibody therapies (11,12) and vaccination (13,14). The distinctive substitutions observed in the spike protein are common, despite the characteristics attributed to some substitutions, for instance, the presence of E484K has been associated with lower neutralizing activity from convalescent plasma(12). The 69/70 deletion spike together with the E484K and N501Y substitutions decrease the ability to neutralize antibodies (15). The insertion 145N in the spike protein is the first evidence in SARS-CoV-2, the implications at the level of infection, transmission and pathogenesis are still unknown.

The B.1.621 lineage includes substitutions targets in RT-PCR screening of VOC lineages (16), this should be considered in the analysis of RT-PCR results considering the occurrence of these mutations in other lineages could lead to overestimating the number of cases caused for VOC lineages.

The B.1.621 lineage was identified in Colombia, USA and Spain. The study was limited to genomic and evolutionary characterization, the public health implications must be to assess through the biological and epidemiologic roles.

References

1. ECDC. SARS-CoV-2 variants of concern as of 4 May 2021 [Internet]. [cited 2021 May 4]. Available from: https://www.ecdc.europa.eu/en/covid-19/variants-concern

2. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 Variants — Clinical, Public Health, and Vaccine Implications. N Engl J Med [Internet]. 2021 Mar 24 [cited 2021 Apr

30];NEJMc2100362. Available from: http://www.nejm.org/doi/10.1056/NEJMc2100362

3. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe [Internet]. 2021 Mar 10 [cited 2021 Apr 30];29(3):463-476.e6. Available from: https://pubmed.ncbi.nlm.nih.gov/33592168/

4. Rees-Spear C, Muir L, Sa G, Heaney J, Aldon Y, Jl S, et al. The impact of Spike mutations on SARS-CoV-2 neutralization 1. bioRxiv [Internet]. 2021 Jan 19 [cited 2021 Mar 2];4:2021.01.15.426849. Available from: https://doi.org/10.1101/2021.01.15.426849

5. Jeyanathan M, Afkhami S, Smaill F, Miller MS, Lichty BD, Xing Z. Immunological considerations for COVID-19 vaccine strategies [Internet]. Vol. 20, Nature Reviews Immunology. Nature Research; 2020 [cited 2021 Apr 30]. p. 615–32. Available from: https://doi.org/10.1038/

6. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science (80-) [Internet]. 2021 Apr 9 [cited 2021 Apr 30];372(6538):eabg3055. Available from: https://science.sciencemag.org/content/372/6538/eabg3055

7. Laiton-Donato K, Villabona-Arenas CJ, Usme-Ciro JA, Franco-Muñoz C, Álvarez-Díaz DA, Villabona-Arenas LS, et al. Genomic epidemiology of severe acute respiratory syndrome coronavirus 2, Colombia. Emerg Infect Dis [Internet]. 2020 Dec 1 [cited 2021 Apr 30];26(12):2854–62. Available from: www.biorxiv.org/content/10.1101/2020.05.21.10932

8. Noticias coronavirus-genoma [Internet]. [cited 2021 Apr 30]. Available from: https://www.ins.gov.co/Noticias/Paginas/coronavirus-genoma.aspx

9. Laiton-Donato K, Usme-Ciro JA, Franco-Muñoz C, Álvarez-Díaz DA, Ruiz-Moreno HA, Reales-González J, et al. Novel highly divergent SARS-CoV-2 lineage with the Spike substitutions L249S and E484K. medRxiv pre-print [Internet]. 2021 Mar 15 [cited 2021 May 7];2021.03.12.21253000. Available from:

https://doi.org/10.1101/2021.03.12.21253000

10. PAHO. Guidance for SARS-CoV-2 samples selection for genomic characterization and surveillance - PAHO/WHO | Pan American Health Organization [Internet]. [cited 2021 May 7]. Available from: https://www.paho.org/en/documents/guidance-sars-cov-2-samples-selection-genomic-characterization-and-surveillance

11. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat Med [Internet]. 2021 Apr 1 [cited 2021 Apr 30];27(4):622–5. Available from: https://doi.org/10.1038/s41591-021-01285-x

12. Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe [Internet]. 2021 Mar 10 [cited 2021 May 1];29(3):477-488.e4. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1931312821000445

13. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature [Internet]. 2021 Feb 10 [cited 2021 Apr 30];592(7855):616–22. Available from: https://www.nature.com/articles/s41586-021-03324-6

14. Garcia-Beltran WF, Lam EC, St. Denis K, Nitido AD, Garcia ZH, Hauser BM, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell [Internet]. 2021 Apr 29 [cited 2021 May 1];184(9):2372-2383.e9. Available from: https://doi.org/10.1016/j.cell.2021.03.013

15. Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med [Internet]. 2021 Apr 1 [cited 2021 May 1];27(4):620–1. Available from: https://doi.org/10.1038/s41591-021-01270-4

16. Bal A, Destras G, Gaymard A, Stefic K, Marlet J, Eymieux S, et al. Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with

spike deletion H69–V70, France, August to December 2020. Eurosurveillance [Internet]. 2021 Jan 21 [cited 2021 May 4];26(3):2100008. Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2021.26.3.2100008

Table1: Nucleotide and amino acid substitution pattern of B.1.621

Genomics coordinate with respect to reference genome (NC_045512.2)	**Amino acid Change**	**Region/Protein**
ORF1ab:		
A3428G	T1055A	Non-structural protein 3
C4878T	T1538I	Non-structural protein 3
A11451G	Q3729R	Non-structural protein 6
C14408T	P4715L	RNA-dependent RNA polymerase
C17491T	P5743S	Helicase domain
ORF3a:		
G25563T	Q57H	accessory protein ORF3a
deletion: 26158-26162 (4 nucleotides)	V256-	accessory protein ORF3a
frameshift	N257X	accessory protein ORF3a
ORF8:		
A27924C	T11K	ORF8 immunoglobulin (Ig) domain protein and related proteins
C28005T	P38S	ORF8 immunoglobulin (Ig) domain protein and related proteins
C28093T	S67F	ORF8 immunoglobulin (Ig) domain protein and related proteins
Spike:		
C21846T	T95I	N-terminal domain of S1 subunit
T21992A, A21993C	Y144T	N-terminal domain of S1 subunit
A21996C, C21997T	Y145S	N-terminal domain of S1 subunit

21998:AAC	ins146N	N-terminal domain of S1 subunit
G22599A	R346K	receptor binding domain
G23012A	E484K	receptor binding domain
A23063T	N501Y	receptor binding domain
C23604A	P681H	S1/S2 cleavage region

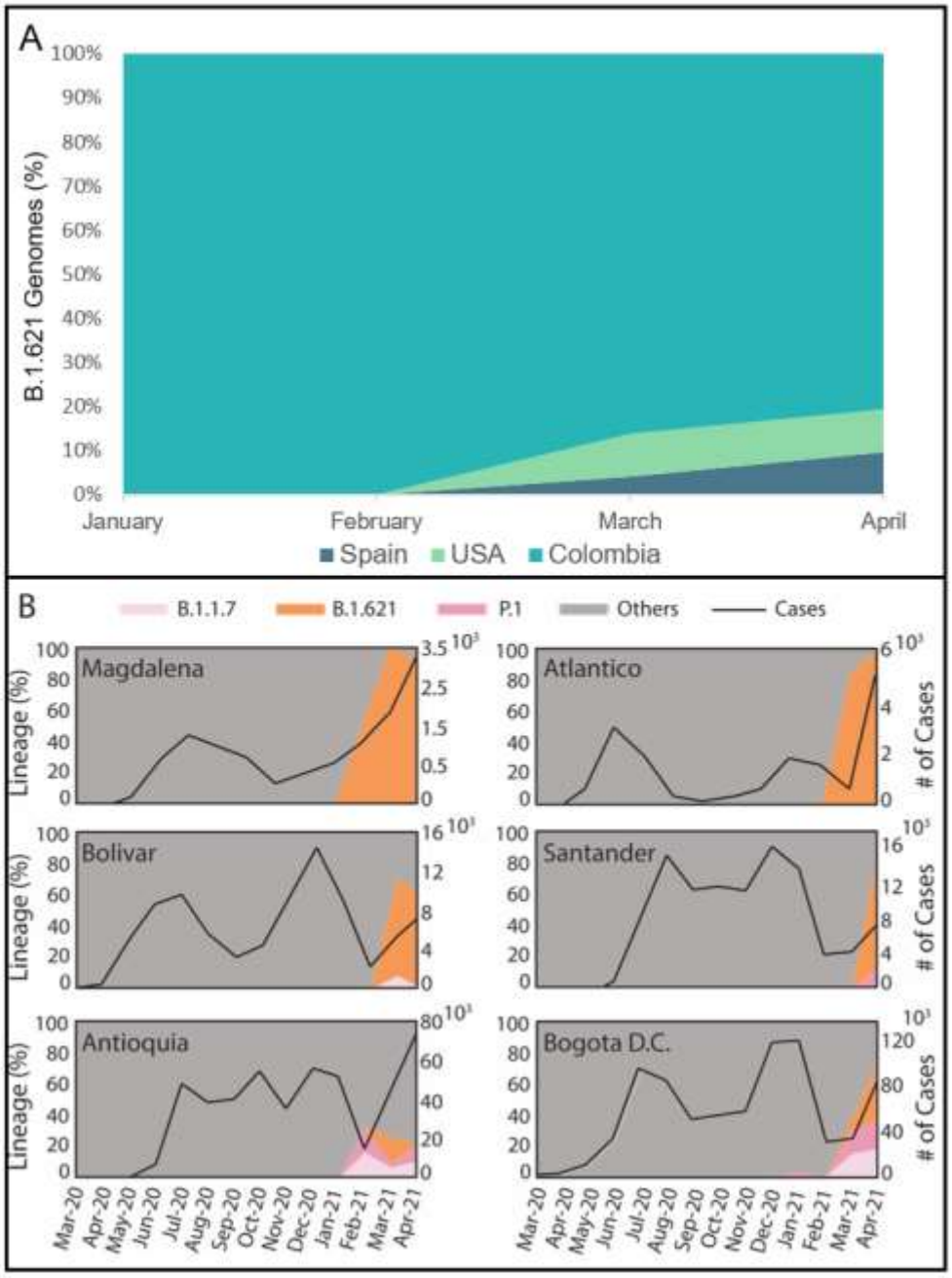


Figure 1. Percent in time of variant B.1.621 in Spain. USA and Colombia.1a) Gisaid registries of variant B.1.621 in 2021. Since January the continuous record has been maintained in Colombia. 1b) Lineage percentage and number of cases of COVID-19 in five departments with circulation of B.1.621 variant and the capital city. B.1.1.7 and P.1 VOC lineages are shown, others lineages circulating are represented as “others”.

Figure 2. Phylogeny and distribution of SARS-CoV-2 b.1.621 variant in Colombia 2a) Phylogenetic tree of the new lineage of SARS-CoV-2 emerging from B.1.621 lineage. The tree was reconstructed by maximum likelihood with the estimated GTR+F+I+G4 nucleotide substitution model for the dataset of 434 genomes. The interactive tree can be accessed in the following link: https://microreact.org/project/5CAiK3qCMaEgE4vYkKVpZW/b7113efc 2b) Map of distribution of lineages across the country

Acknowledgements

The authors thank the National Laboratory Network for routine virologic surveillance of SARS-CoV-2 in Colombia. We also thank all researchers who deposited genomes in GISAID's EpiCoV Database contributing to genomic diversity and phylogenetic relationship of SARS-CoV-2. We thank Rotary International and Charlie Rut Castro for equipment's donation. Finally, we thank red RENATA and Universidad Industrial de Santander for the workstation bioinformatic support.

Disclosure statement

No conflict of interest was reported by the authors.

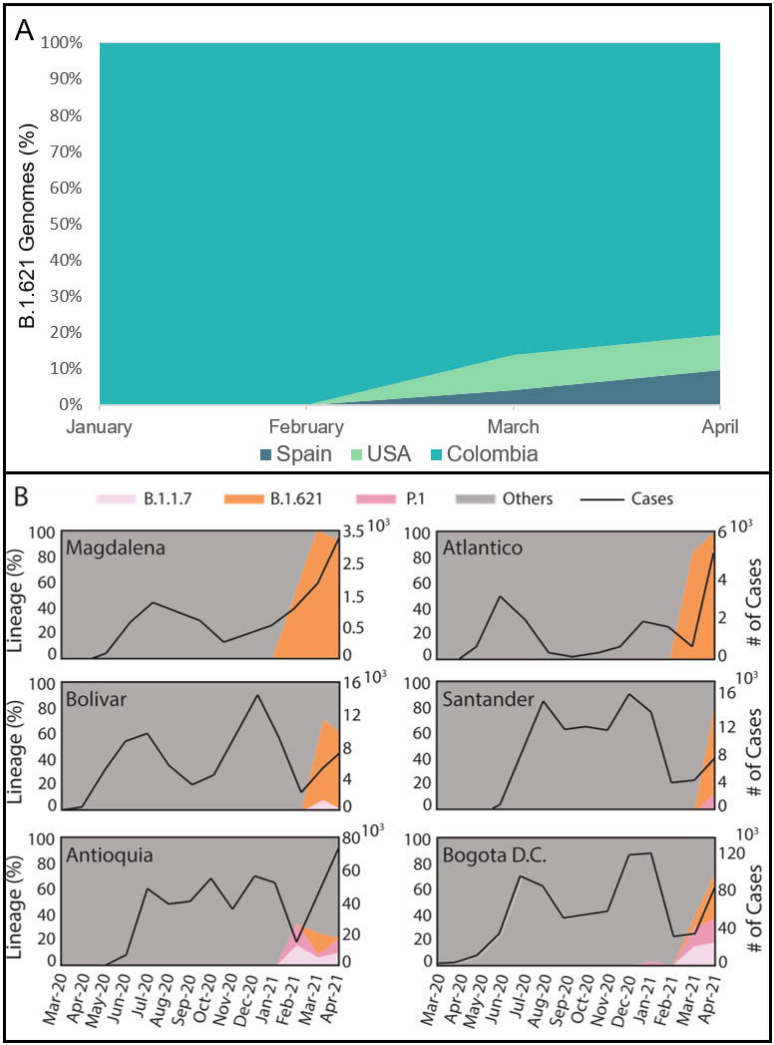
Funding

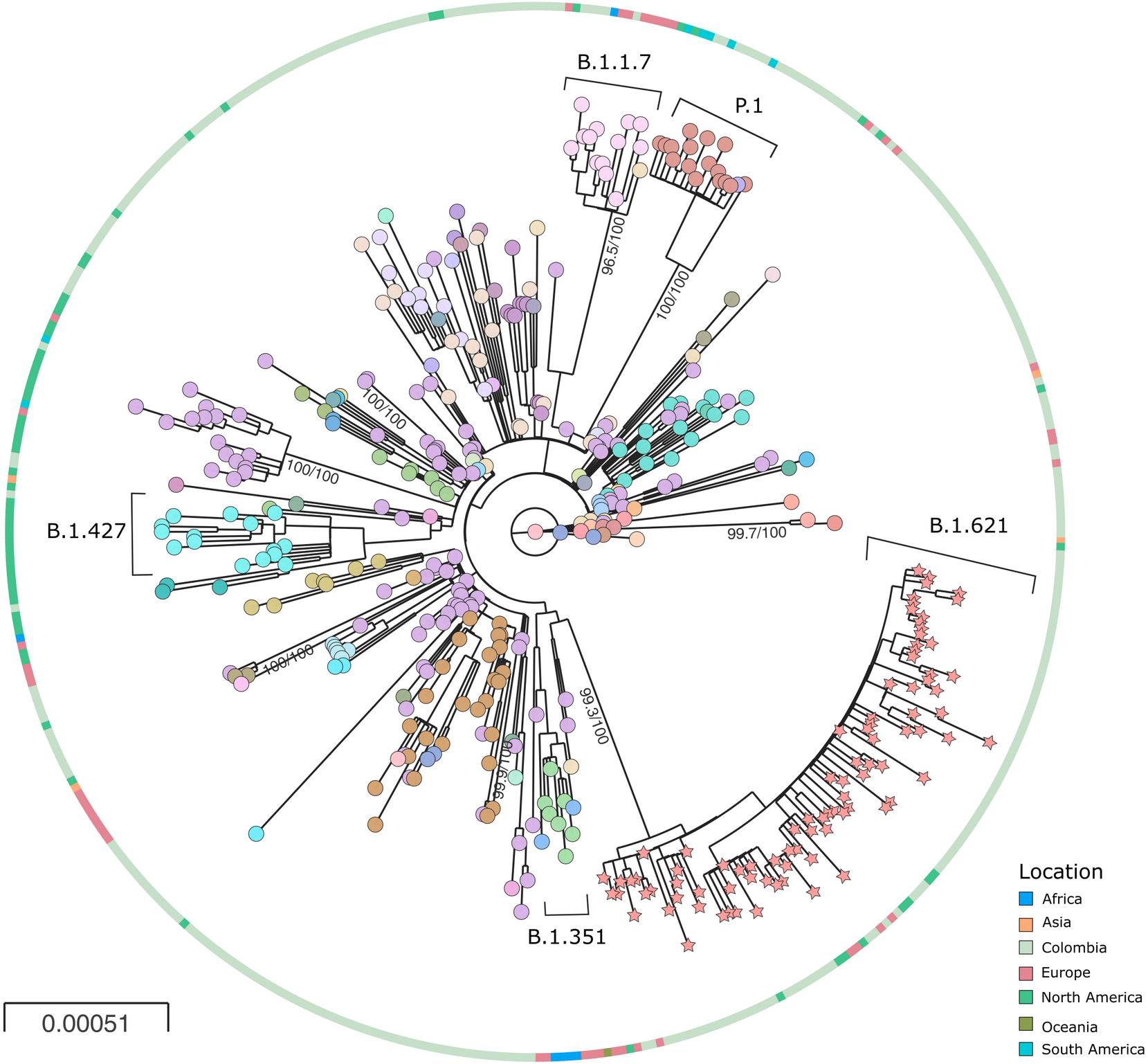
This work was funded by the Project CEMIN-4-2020 Instituto Nacional de Salud. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

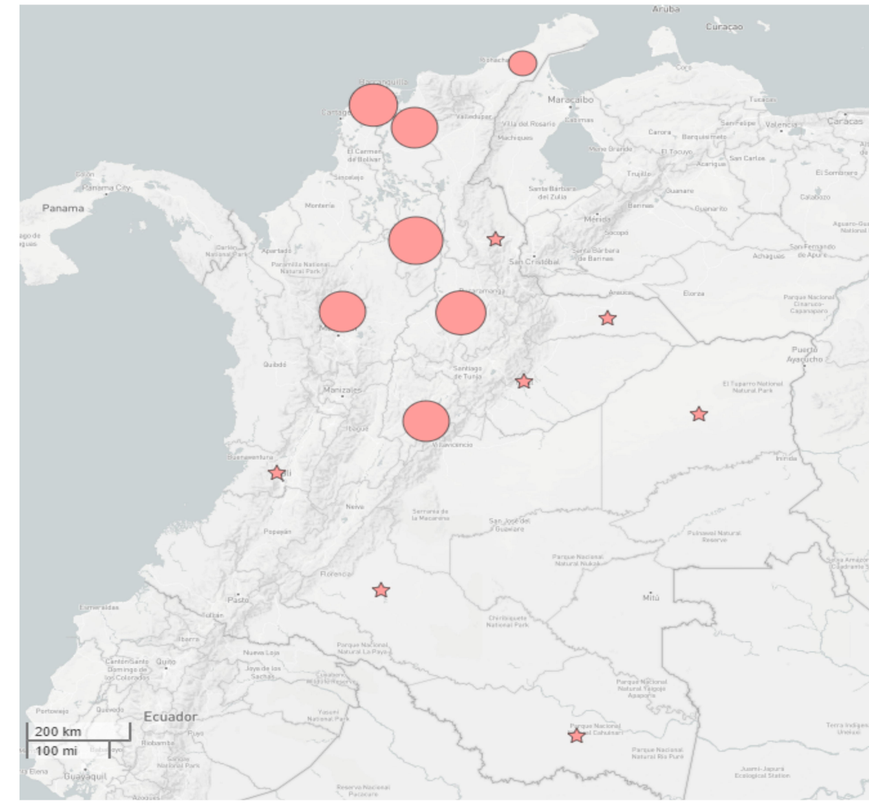
Data deposition

SARS-CoV-2 Colombian sequences belonging to the B.1.621 were deposited in GISAID under accession numbers: EPI_ISL_1220045, EPI_ISL_1582980, EPI_ISL_1424054, EPI_ISL_1424056, EPI_ISL_1582978, EPI_ISL_1820926, EPI_ISL_1424055, EPI_ISL_1582993, EPI_ISL_1582979, EPI_ISL_1424057, EPI_ISL_1820929, EPI_ISL_1820930, EPI_ISL_1820932, EPI_ISL_1820927, EPI_ISL_1424058, EPI_ISL_1582984, EPI_ISL_1582986, EPI_ISL_1582991, EPI_ISL_1820958, EPI_ISL_1582990, EPI_ISL_1582992, EPI_ISL_1582981, EPI_ISL_1582994, EPI_ISL_1582988, EPI_ISL_1820959, EPI_ISL_1820928, EPI_ISL_1582996, EPI_ISL_1632530, EPI_ISL_1582989, EPI_ISL_1820934, EPI_ISL_1582987, EPI_ISL_1820955, EPI_ISL_1820935, EPI_ISL_1582997, EPI_ISL_1820925, EPI_ISL_1820950, EPI_ISL_1820949, EPI_ISL_1820944, EPI_ISL_1820947, EPI_ISL_1820943, EPI_ISL_1820946, EPI_ISL_1820954, EPI_ISL_1821882, EPI_ISL_1820939, EPI_ISL_1820942, EPI_ISL_1820960, EPI_ISL_1820936, EPI_ISL_1820948, EPI_ISL_1820931, EPI_ISL_1582977, EPI_ISL_1820965, EPI_ISL_1821070, EPI_ISL_1821075, EPI_ISL_1821063, EPI_ISL_1820967, EPI_ISL_1820968, EPI_ISL_1821062, EPI_ISL_1821064, EPI_ISL_1821069, EPI_ISL_1821073, EPI_ISL_1821074, EPI_ISL_1821076, EPI_ISL_1821077, EPI_ISL_1821071, EPI_ISL_1821072, EPI_ISL_1820962, EPI_ISL_1821065,

EPI_ISL_1821066, EPI_ISL_1821067, EPI_ISL_1821068, EPI_ISL_1824702, EPI_ISL_1824703, EPI_ISL_1824706, EPI_ISL_1824711, EPI_ISL_1824712, EPI_ISL_1824713, EPI_ISL_1824714.


A
B.1.621 Genomes (%)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
January
February
March
April
Spain
USA
Colombia
B
B.1.1.7
B.1.621
P.1
Others
Cases
Magdalena
Atlantico
Bolivar
Santander
Antioquia
Bogota D.C.
Lineage (%)
# of Cases
Mar-20
Apr-20
May-20
Jun-20
Jul-20
Aug-20
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21


B.1.1.7
P.1
96.5/100
100/100
100/100
100/100
B.1.427
99.7/100
B.1.621
100/100
99.3/100
99.6/100
B.1.351
0.00051
Location
Africa
Asia
Colombia
Europe
North America
Oceania
South America


Aruba
Curaçao
Maracaibo
Caracas
Panama
Ecuador
200 km
100 mi