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


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.
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


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

<table>
<thead>
<tr>
<th>Genomics coordinate with respect to reference genome (NC_045512.2)</th>
<th>Amino acid Change</th>
<th>Region/Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORF1ab:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3428G</td>
<td>T1055A</td>
<td>Non-structural protein 3</td>
</tr>
<tr>
<td>C4878T</td>
<td>T1538I</td>
<td>Non-structural protein 3</td>
</tr>
<tr>
<td>A11451G</td>
<td>Q3729R</td>
<td>Non-structural protein 6</td>
</tr>
<tr>
<td>C14408T</td>
<td>P4715L</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>C17491T</td>
<td>P5743S</td>
<td>Helicase domain</td>
</tr>
<tr>
<td><strong>ORF3a:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G25563T</td>
<td>Q57H</td>
<td>accessory protein ORF3a</td>
</tr>
<tr>
<td>deletion: 26158-26162 (4 nucleotides)</td>
<td>V256-</td>
<td>accessory protein ORF3a</td>
</tr>
<tr>
<td>frameshift</td>
<td>N257X</td>
<td>accessory protein ORF3a</td>
</tr>
<tr>
<td><strong>ORF8:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A27924C</td>
<td>T11K</td>
<td>ORF8 immunoglobulin (Ig) domain protein and related proteins</td>
</tr>
<tr>
<td>C28005T</td>
<td>P38S</td>
<td>ORF8 immunoglobulin (Ig) domain protein and related proteins</td>
</tr>
<tr>
<td>C28093T</td>
<td>S67F</td>
<td>ORF8 immunoglobulin (Ig) domain protein and related proteins</td>
</tr>
<tr>
<td><strong>Spike:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21846T</td>
<td>T95I</td>
<td>N-terminal domain of S1 subunit</td>
</tr>
<tr>
<td>T21992A, A21993C</td>
<td>Y144T</td>
<td>N-terminal domain of S1 subunit</td>
</tr>
<tr>
<td>A21996C, C21997T</td>
<td>Y145S</td>
<td>N-terminal domain of S1 subunit</td>
</tr>
<tr>
<td>Variation</td>
<td>Change</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>21998:AAC</td>
<td>ins146N</td>
<td>N-terminal domain of S1 subunit</td>
</tr>
<tr>
<td>G22599A</td>
<td>R346K</td>
<td>receptor binding domain</td>
</tr>
<tr>
<td>G23012A</td>
<td>E484K</td>
<td>receptor binding domain</td>
</tr>
<tr>
<td>A23063T</td>
<td>N501Y</td>
<td>receptor binding domain</td>
</tr>
<tr>
<td>C23604A</td>
<td>P681H</td>
<td>S1/S2 cleavage region</td>
</tr>
</tbody>
</table>
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

[Phylogenetic tree image]

[Distribution map image]

[Interactive tree link]
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_1582993, EPI_ISL_1820932, EPI_ISL_1820927, EPI_ISL_1424057, EPI_ISL_1820926, EPI_ISL_1424055, EPI_ISL_1582984, EPI_ISL_1582986, EPI_ISL_1582991, EPI_ISL_1820927, EPI_ISL_1582990, EPI_ISL_1582992, 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_1582997, EPI_ISL_1820925, EPI_ISL_1820950, EPI_ISL_1820949, EPI_ISL_1820944, EPI_ISL_1820947, EPI_ISL_1820943, EPI_ISL_1820949, EPI_ISL_1820944, EPI_ISL_1820947, EPI_ISL_1820943, EPI_ISL_1820942, EPI_ISL_1820960, EPI_ISL_1820936, EPI_ISL_1820948, EPI_ISL_1820935, EPI_ISL_1582997, EPI_ISL_1820925, 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.