TITLE: MULTIPLE INDEPENDENT EVOLUTIVE EVENTS RELATED TO SARS-COV-2 P.1 (GAMMA) VARIANT OF CONCERN ACROSS BRAZIL

AUTHORS: LOVISON, O.A.; WINK, P.L.; RAMALHO, R.G.; MONTEIRO, F.; VOLPATTO, F.; WILLIG, J.; ZAVASCKI, A.P.; BARTH, A.L..; MARTINS, A.F.

INSTITUTION: UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, PORTO ALEGRE, RS.

ABSTRACT:

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) variant of concern (VOC) P.1 (Gamma) emerged in Brazil in December 2020 in Amazonas (AM) state and in February 2021 in Rio Grande do Sul (RS) state. However, previous studies reported P.1 cases in RS in November 2020, three months before the increase in case numbers and hospitalizations. To better understand the emergence of P.1 lineage, we built a dataset of SARS-CoV-2 sequences as follows: A) 222 sequences of B.1.1.28 lineage from AM (n=49) and RS (n=173) collected from July 2020 to February 2021; B) 164 sequences of P.1 lineage from AM (n=127) and RS (n=37) collected up to February 2021; C) The six oldest P.1 sequences from other Brazilian states, collected until November, 2020. The dataset "C" was used to calibrate the molecular clock. The sequences were downloaded from GISAID. The alignment was performed by MAFFT v7.475; the Bayesian analysis was performed using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v2.6.5 with a strict molecular clock; GTR + F + I + G4 model, and a non-parametric Bayesian skyline (BSKL) model as the coalescent tree prior. Five independent runs were performed with 150 million MCMC chains. Convergence and mixing were performed using TRACER v1.7.2. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v2.6.4. All trees built in this study were visualized using FigTree v1.4.4 and designed using R v4.1.0 packages. The molecular clock demonstrated that P.1 sequences were grouped into a single well-supported clade (posterior probability = 0.99) with a common ancestor dating of around May 28, 2020 (95% HPD: 2020-01-24 to 2020-06-26). The common ancestor of the 164_LABRESIS (EPI_ISL_3233232) sequence dated of around June 20, 2020 (95% HPD: 2020-04-28 to 2020-08-26), and the RS-LMM38991 (EPI_ISL_2241496) sequence was from around July 25, 2020 (95% HPD: 2020-06-08 to 2020-08-04). These sequences are the oldest P.1 sequences from RS and there is no epidemiological link to AM sequences. The time-scale demonstrates that the P.1 sequences from AM have an estimated date of origin prior to the ones from RS. However, the dataset C (common ancestor dating around July 10) showed a divergence time higher than the oldest sequence from AM (dating around November 24). These results indicate that the P.1 lineage had multiple introduction events in the country and could have emerged in different regions of Brazil before the first outbreak in AM.

Keywords: SARS-CoV-2, Variant of Concern P.1, Gamma