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.
Research paper

E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil

Patrícia Aline Grôhs Ferrareze, Vinicius Bonetti Franceschi, Amanda de Menezes Mayer, Gabriel Dickin Caldana, Ricardo Ariel Zimerman, Claudia Elizabeth Thompson

A Graduate Program in Health Sciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil
b Center of Biotechnology, Graduate Program in Cell and Molecular Biology (PPGBCM), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
c Irmãando Santa Casa de Misericórdia de Porto Alegre, Porto Alegre, RS, Brazil
d Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

ARTICLE INFO

Keywords:
COVID-19
E484K
Severe acute respiratory syndrome coronavirus 2
Infectious diseases
Viral evolution

ABSTRACT

The COVID-19 pandemic caused by SARS-CoV-2 has affected millions of people since its beginning in 2019. The propagation of new lineages and the discovery of key mechanisms adopted by the virus to overlap the immune system are central topics for the entire public health policies, research and disease management. Since the second semester of 2020, the mutation E484K has been progressively found in the Brazilian territory, composing different lineages over time. It brought multiple concerns related to the risk of reinfection and the effectiveness of new preventive and treatment strategies due to the possibility of escaping from neutralizing antibodies. To better characterize the current scenario we performed genomic and phylogenetic analyses of the E484K mutated genomes sequenced from Brazilian samples in 2020. From October 2020, more than 40% of the sequenced genomes present the E484K mutation, which was identified in three different lineages (P.1, P.2 and B.1.1.33 - posteriorly renamed as N.9) in four Brazilian regions. We also evaluated the presence of E484K associated mutations and identified selective pressures acting on the spike protein, leading us to some insights about adaptive and purifying selection driving the virus evolution.

1. Introduction

The etiological agent of COVID-19, named SARS-CoV-2, belongs to the Coronaviridae family composed of enveloped positively oriented single-stranded RNA viruses (Zhu et al., 2020). The first cases of the disease were reported in Wuhan province, China, in December 2019. Since then the number of infected people has increased exponentially worldwide and the World Health Organization (WHO) declared it a pandemic on 11 March 2020. As of 24 January 2021, the number of confirmed cases globally has scaled up to 99 million, with over 2 million deaths. Brazil is the third country in total number of confirmed cases, already exceeding 8.8 million infected people and 216,445 deaths (Johns Hopkins Coronavirus Resource Center, 2021). Compared to Latin America, which represents ~18% of the world cases, Brazil accounts for almost half of the COVID-19 notifications. However, the number of cases per million people (~41,600) is comparable to other countries such as Argentina (~41,300), Colombia (~39,600), Costa Rica (~37,160), Chile (~36,570), and Peru (~33,170) (Ritchie et al., 2021).

Nearly 410,000 genomes (up to January 2021) sequenced by researchers from several countries are available on the Global Initiative on Sharing All Influenza Data (GISAID) (Shu and McCauley, 2017), which is crucial to investigate the SARS-CoV-2 genomic epidemiology. A standardized nomenclature was proposed to reflect genetic characteristics and the viral geographical spread patterns. Two ancestral lineages, A and B, have been described containing sublineages that are distinguished by different recurrent mutations and phylogenetic profiles (Rambaut et al., 2020a). More recently, special attention has been directed to the understanding of aspects that might impact the virulence and transmissibility of SARS-CoV-2 (Korber et al., 2020; Li, 2020; Toyoshima et al., 2020; Volz et al., 2020) and the influence of different mutations in the effectiveness of COVID-19 vaccines and immunotherapies (Andreano et al., 2020; Pinto et al., 2020; Yu et al., 2020).

Some viral lineages have been studied in more detail, mainly those carrying mutations in the spike (S) glycoprotein, since it is the binding...
Infection, Genetics and Evolution 93 (2021) 104941

In January 2021, the E484K mutation was identified in several viral genomes from Brazil. Due to its rapid spread, the independent origins and the potential implications in vaccination and passive immune therapies, E484K has received particular attention ever since. In Qatar, 14 days after the second vaccine dose (Pfizer–BioNTech), the effectiveness against the B.1.351 variant was 75.0% (Abu-Raddad et al., 2021); in South Africa, the AstraZeneca ChAdOx1 vaccine presented only 10% protection against mild-to-moderate disease associated with the B.1.351, while 75% of protection was obtained against B.1.1.7 in the U.K. (Gupta, 2021). However, in Brazil, the CoronaVac/Sinovac effectiveness was estimated at around 50% against symptomatic infection P.1 variant (Gupta, 2021). Importantly, B.1.351, P.1, and P.2 carry this substitution associated with escape from neutralizing antibodies (Baum et al., 2020; Greaney et al., 2021; Weisblum et al., 2020). Recently, this mutation was also identified in a sample from a reinfected patient (Nonaka et al., 2021).

Moreover, B.1.1.7, B.1.351, and P.1 also harbor N501Y mutation, which is associated with enhanced receptor affinity (Starr et al., 2020) and increased infectivity and virulence in a mouse model (Gu et al., 2020). The presence of E484K and N501Y substitutions in the same SARS-CoV-2 genome may be particularly relevant for viral evolution. This combination was shown to induce more conformational changes than the N501Y mutant alone, potentially altering antibodies’ complementarity to this region resulting in the above mentioned immune escape phenomena (Nelson et al., 2021).

Structural analysis points to E484K as potentially the most crucial mutation so far in brazilian SARS-CoV-2 genomes. It creates a new site for the amino acid 75 hACE-2 binding. This interaction seems even
Table 1
Lineage-defining mutations of each of the three Brazilian lineages carrying the E484K mutation. These mutations do not necessarily reflect pangolin lineage assignment defining-mutations, but were extracted based on their representatives in the majority of sequences of each lineage (https://github.com/cov-vlines/pangolin).

| Lineage/ | Mutation | N.9 (B.1.1.33 with E484K) | P.1 | P.2 |
|----------|----------|---------------------------|-----|-----|
| 5’ UTR   | ORF1ab   |                           |     |     |
|          | G1264T (Nsp2) | G1264T (Nsp2)           |     |     |
|          | C5573T (Nsp2) | C5573T (Nsp2)          |     |     |
|          | S2103F (Nsp2) | S2103F (Nsp2)         |     |     |
|          | C7600T (Nsp3) | C7600T (Nsp3)         |     |     |
|          | A2529V (Nsp3) | A2529V (Nsp3)         |     |     |
|          | T11078C (Nsp6) | T11078C (Nsp6)      |     |     |
|          |                          | G11288-11,296 (Nsp6) |     |     |
|          |                          | F36065L (3675-3677 SF) |     |     |
|          |                          | C11277T (Nsp9)        |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
|          |                          |                          |     |     |
The N.9 (B.1.1.33.9) lineage carrying the E484K mutation was found in São Paulo (n = 3) and Amazonas (n = 1), while P.1 sequences were found only in the Amazonas state (n = 19). All other sequences (n = 111) belong to the P.2 lineage, the most widespread lineage considering sequenced data available so far. The Northeast region was represented by sequences from Alagoas (n = 2), Paraiba (n = 3), and Bahia (n = 1). The North region is represented by Amazonas sequences (n = 6). Southeast sequenced 44 P.2 genomes, 36 from Rio de Janeiro and 8 from São Paulo. Southern Brazil classified 55 genomes as P.2, 5 from Parana and 50 from Rio Grande do Sul, reinforcing its probable dissemination to all regions of Brazil (Fig. 3).

The worldwide emergence of E484K began in March 2020, with three sequences firstly represented. A significant increase was observed in October (n = 86), followed by successive increases in November (n = 366) and December (n = 374) (Fig. 4A). A similar trend was observed in Brazil, as E484K was observed in sequences obtained from October (n = 31), November (n = 87), and December (n = 40). Most importantly, the proportion of genomes carrying this replacement was 39.7%, 43.9%, and 43.5% from October to December in comparison to all genomes sequenced from Brazil (Fig. 4B). We believe this apparent slow-growing pattern in Brazil is due to heterogeneous and limited initiatives of sequencing in the country, and probably this substitution is already widespread through Brazilian states, as its harbored by three distinct and apparently independent evolving lineages.

2.2. Selection analysis

In order to obtain a reliable detection of sites submitted to positive/adaptive or negative/purifying selection, a random set of Brazilian genomes was tested with different approaches. Regarding individual site models, the Bayesian inference FUBAR (Fast, Unconstrained Bayesian Approximation) identified eight sites evolving under adaptive selection (positive selection) in the spike sequence (Table 2), with calculated synonymous and nonsynonymous average rates of 1.227 and 0.857, respectively. For these residues under adaptive pressure, six are included in known mutation sites of spike protein, including E484K (L5F, S12F, P26S, D138Y, A688V).

The analysis with the codeml program from the PAML package confirmed all the positively selected sites predicted by FUBAR. The M3 vs. M0 (p < 0.001) comparison indicated a highly variable selective pressure among sites, while the M2 vs. M1 and M8 vs. M8a models comparison (p < 0.05) indicated some sites submitted to positive selection (Tables 3, 4 and 5). The results for the M3 model indicated a kappa value (ts/tv) of 2.82630 and a site proportion of 93.156% with an estimated omega (ω) value of 0.25424 (negative selection), while 3.536% of the sites were estimated to have ω = 4.90450 (positive selection) (Table 3). Differently from M2 that is compared with M1 to detect sites under positive selection, the M3 model for discrete evolution also indicates sites under a neutral selection (p1, ω1), being more recommended to evaluate variable selective pressures among sites. The proportion of sites and the ω values obtained for positive selection by M2 and M3 models were almost equal (M2: ω = 4.90459). In the case of M7 vs. M8 and M8a vs. M8 comparisons, the rejection of the alternative model (M8) by the absence of a significant likelihood ratio suggests a beta distribution of the ω. When applied to the analysis of the spike protein, the results of M8a, which is a more refined and conservative model, allows us to confirm that some residues of this protein are actually under adaptive selection.

The ω estimator (ω = dS/dD) is a common method to detect positive selection, since it assumes that synonymous substitutions are neutral.

Fig. 2. Bayesian phylogenetic inference of the 134 Brazilian E484K mutated genomes. Tips were colored by Brazilian state and the reference genome NC_045512.2 is represented in black. The branches were highlighted by lineage: green (B.1.1.33), light green (N.9), beige (B.1.1.28), light red (P.2) and blue (P.1). Mutations occurring in all analyzed genomes for each lineage were described next to the respective nodes. The asterisks indicate the SARS-CoV-2 genomes sequenced by our research group.
and the nonsynonymous are subject to selection. Consequently, a ω statistically higher than 1 would indicate the action of positive selection or a relaxed selective constraint, whereas low dS/dω values would mean conservation of the gene product due to purifying selection (Tennesen, 2008). Additionally to the FUBAR results, three sites were also considered with ω > 1 (Tables 3 and 5): 222, 626, and 1263. Of these, only site 222 is not related to the E484K-presenting lineages (A626S and P1263L/S are known). The list of positively selected sites identified by the M8 model is available on Table S4.

To avoid the potential missense effects of the function lost by the sequence mutation, the purifying selection acts as a protective model. The FEL (Fixed Effects Likelihood) approach detected 19 sites under negative selection (Table 6). Three of them (189, 191, and 564) are near residues presenting known nonsynonymous mutations in the E484K mutated genomes, as S:R190S and S:F565L. The SLAC (Single-Likelihood Ancestor Counting) method identified four sites (55, 856, 943, and 222) as S:R190S and S:F565L. The list of positively selected sites identified by the M8 model is available on Table S4.

In summary, we have demonstrated widespread dissemination of E484K mutation across different lineages (A) and Brazilian states (B) from October to December 2020, and the nonsynonymous are subject to selection. Consequently, a ω statistically higher than 1 would indicate the action of positive selection or a relaxed selective constraint, whereas low dS/dω values would mean conservation of the gene product due to purifying selection (Tennesen, 2008). Additionally to the FUBAR results, three sites were also considered with ω > 1 (Tables 3 and 5): 222, 626, and 1263. Of these, only site 222 is not related to the E484K-presenting lineages (A626S and P1263L/S are known). The list of positively selected sites identified by the M8 model is available on Table S4.

To avoid the potential missense effects of the function lost by the sequence mutation, the purifying selection acts as a protective model. The FEL (Fixed Effects Likelihood) approach detected 19 sites under negative selection (Table 6). Three of them (189, 191, and 564) are near residues presenting known nonsynonymous mutations in the E484K mutated genomes, as S:R190S and S:F565L. The SLAC (Single-Likelihood Ancestor Counting) method identified four sites (55, 856, 943, and 1215) of which were already predicted by FEL. The prevalence of negative selection across the spike protein is consistent with the low genome-wide mutation rate inferred for SARS-CoV-2 (van Dorp et al., 2020).

3. Discussion

The permanence of a pathogen inside a population host depends on its efficiency in key processes such as the replication, the escaping of the immune system and the binding to the cell receptor. Mutations that create an advantageous scenario towards the host response to infection enhance the pathogen fitness under the natural selection pressure. Consequently, the host-pathogen coevolution represents an important mechanism to understand the establishment and the prognostics of pathogenesis, since the infections are possibly the major selective pressure acting on humans (Sironi et al., 2015). Therefore, the circulation of low and moderate pathogens provides time for the pathogen adaptations, in such a way that the modulation of the immunity may possibly promote molecular convergence in different lineages over the time (Longdon et al., 2014).

The presence of spike S:D614G, N:R203K, N:G204R, and Nsp12-P323L in all sequenced E484 mutated samples, reaching three different lineages, might suggest that these lineages show increased viral replication. Considering that E484K enhances escape from immune system antibodies, these may potentially lead to a viral advantage. The occurrence of simultaneous mutations as N:R203K and N:G204R is already known in the SARS-CoV-2 literature. However, the fixation of these mutations, as well as Nsp12-P323L and D614G in all the E484K evaluated genomes may indicate a novel adaptive relationship among these modifications resulting in viral evolutionary success.

Even as an independent evolutionary event, the potential fixation of mutations such as E484K across lineages may indicate active mechanisms of adaptive selection and are very relevant in planning future therapeutic strategies (for example, newer vaccines and immunotherapy platforms). The deletion of three amino acids in the second helix of the transmembrane protein Nsp6 in P.1 genomes may affect the virus-induced cellular autophagy and the formation of double-membrane vesicles for the viral RNA synthesis (Benvenuto et al., 2020).

The important role of subsequent stabilization of the flexible NTD by mutations has been speculated (Laha et al., 2020). Typically, NTD can harbor a larger number of evolutionary events than RBD, including mutations, insertions and deletions that could act allosterically altering the binding affinities between RBD and hACE-2 and inducing immune evasion. It is known that distinct positions may have a linked relationship in the final protein structure and may display some advantage by acting together to achieve increased stability, adaptability, viability and/or transmission efficiency (Laha et al., 2020). The potential causality or influence of the E484K and other substitutions for the effectiveness of neutralizing antibodies that bind the N-terminal domain of the spike protein (Chi et al., 2020) remains uncertain.

Concerning the selection analysis, the FUBAR method was the first one to be tested. According to Murrell et al. (2013), the FUBAR method may have more detection power than methods like FEL, in particular when positive selection is present but relatively weak. The analysis of the South African clade V501V2 (Pond et al., 2020) found some similar results for FUBAR evaluation, with the detection of adaptive selection at the sites 5 (L5F), 12 (S12F), and 484 (E484K). The residues 155 and 677 are not associated with the E484K-presenting lineages, however, recent data found evidence of evolutionary convergence of this mutation in at least six distinct sub-lineages, which could improve proteolytic processing, cell tropism, and transmissibility (Hodcroft et al., 2021). The amino acid residue 677 is next to the furin-like cleavage site (678 to 688). Interestingly, some studies suggest that the emergence of the SARS-CoV-2 in human species resulted from a five amino acid change in the critical S glycoprotein binding site (Fam et al., 2020).

To the best of our knowledge, the impact of E484K in different lineages has not been deeply explored. It was structurally demonstrated that, at least in combination with K417N and N501Y, the substitution has profound impact in shifting the main site of contact between viral RBD and hACE-2 residues (Nelson et al., 2021). However, we still do not know if this holds true for other sets of mutations associated with E484K.

In summary, we have demonstrated widespread dissemination of mutants harboring E484K replacement in geographically diverse regions in Brazil, as well as the potential fixation of the E484K mutation despite a short time (three months) after its first arising. This substitution was disseminated in our country as early as October 2020, although phylogenetic analyses estimated that indels were fixed in Brazil and in the P.1 ancestor by July 2020. The fact that E484K was found in the context of different mutations and lineages is suggestive of the particular substitution may act as a common solution for viral evolution in different genotypes. This hypothesis may be related to the profound impact of the mutation, which changes a negatively charged amino acid (glutamic acid) for a positively charged amino acid (lysine). Since this position is present in a highly
flexible loop, it has been proposed that the presence of such mutation could create a strong ionic interaction between lysine (K) in RBD and amino acid 75 of the hACE-2 receptor shifting the major sites of binding in S1 from positions 497–502 to 484 (Nelson et al., 2021).

Mutations in RBD could theoretically decrease the neutralizing serum activity of patients receiving vaccines against SARS-CoV-2 with unknown clinical consequences. Arguably the effect of E484K could be particularly relevant. In a previous publication, this mutation was associated with complete abolishment of all neutralizing activity in a high proportion of convalescent serum tested (Wibmer et al., 2021).

Jangra et al. (2021) showed that the E484K affects the binding of serum polyclonal neutralizing antibodies and decreases the neutralization efficiency for serum with low or moderate IgG for the SARS-CoV-2 spike protein. Moreover, in the study of Xie et al. (2021), the combination of the mutations E484K + N501Y + D614G generated lower neutralization titers than the N501Y virus or the virus with three mutations from the UK variant (Δ69/70 + N501Y + D614G). When taken together, this growing body of evidence suggests that E484K should be the target of intense virologic surveillance. Studies testing the activity of serum from vaccinated patients against viruses or pseudoviruses with the aforementioned substitution should be considered a high public health priority. Second generation immune therapies and vaccines focusing on more conserved domains (for instance, in S2 fusion domain) may deserve special attention to assure continuous therapeutic efficacy.

4. Methods

4.1. Sequencing data retrieval

For the graphic evaluation of Brazilian lineages, all genomes available on the GISAID database until January 18th, 2021, and presenting the E484K mutation were included in the analysis. Genomes from Brazil and other countries submitted until January 24th, 2021, and with collected dates between January 1st and December 31st, 2020, were selected. The counting of genomes containing E484K was performed by the specification of the mutation in the search fields. The monthly count was defined by the values between the first and the last day of each month.

4.2. Comparative genomic analyses

The comparative genomic analysis of the E484K mutated genomes...
Infection, Genetics and Evolution 93 (2021) 104941

and M8a variable E484K mutation for Brazil in GISAID until January 18th, 2021. For was performed with all the 134 genomes (Table S1) presenting the -sites (*: M2 selection model: Naive Empirical Bayes (NEB) analysis. Positively selected PAML (codeml) site table for positively selected sites. Table 5

| Model | Parameters | InL | Sites showing indications of positive selection |
|-------|------------|-----|-----------------------------------------------|
| M0    | ω = 0.40548 | -6617.469118 | None |
| M1    | p0 = 0.64290, p1 = 0.35710 | -6607.96566 | Not allowed |
| M2    | p0 = 0.96461, p1 = 0.00003, p2 = 0.03536, u0 = 0.25444, u1 = 1, u2 = 4.90459 | -6604.579083 | 5, 12, 26, 138, 155, 222, 484, 626, 677, 688, 1263 |
| M3    | p0 = 0.93156, p1 = 0.03308, p2 = 0.03536, u0 = 0.25424, u1 = 0.25458, u2 = 4.90450 | -6604.579553 | 5, 12, 26, 138, 155, 222, 484, 626, 677, 688, 1263 |
| M4    | p0 = 0.95914, q0 = 0.00475 | -6605.852845 | 5, 12, 14, 20, 26, 27, 52, 54, 68, 138, 145, 153, 155, 190, 218, 221, 222, 223, 231, 235, 263, 344, 417, 658, 670, 677, 688, 776, 791, 879, 936, 1005, 1065, 1071, 1072, 1076, 1099, 1104, 1118, 1152, 1162, 1238, 1259, 1263, 1264, 1272 |
| M5    | p = 0.00986, q = 0.01846 | -6608.318528 | None |
| M6    | p0 = 0.91883, q0 = 0.04574, p1 = 0.08117, | -6607.965703 | 5, 12, 26, 138, 155, 222, 484, 626, 677, 688, 1263 |
| M7    | ω = 2.79061 | -6617.469118 | None |

ω = δω/δq = average over sites; p0,p1 and p2 indicate the proportions of groups 0, 1 and 2 in each model, respectively; u0, u1 and u2 indicate the ω values of groups 0, 1 and 2 in each model, respectively. p and q are beta parameters.

Table 4

| Comparison | 2Δ/ ) | Probability values (p) |
|------------|-------|------------------------|
| M3 vs. M0  | 25.78033 | <0.001 |
| M2 vs. M1  | 6.773154 | <0.05 |
| M8 vs. M7  | 4.903166 | <0.1 |
| M8 vs. M8a | 4.225314 | <0.1 |

The degrees of freedom used comparing models M3 vs. M0, M2 vs. M1, M8 vs. M7 and M8a vs. M8 are 4, 2, 2 and 1, respectively. Probability values ≤ 0.05 are considered as statistically significant.

Table 5

| Site | NEB probabilities | BEB probabilities |
|------|-------------------|-------------------|
|      | Prob ( ω > 1) | post mean for ω | Prob ( ω > 1) | post mean ± SE for ω |
| 5 L  | 0.988* | 4.850 | 0.892 | 2.153 ± 0.516 |
| 12 S | 0.988* | 4.849 | 0.892 | 2.153 ± 0.517 |
| 26 P | 0.989* | 4.854 | 0.894 | 2.156 ± 0.514 |
| 138 D | 0.758 | 3.779 | 0.741 | 1.922 ± 0.719 |
| 155 S | 0.782 | 3.893 | 0.749 | 1.935 ± 0.712 |
| 222 A | 0.775 | 3.857 | 0.746 | 1.930 ± 0.714 |
| 484 E | 0.841 | 4.163 | 0.770 | 1.968 ± 0.692 |
| 626 A | 0.860 | 4.252 | 0.778 | 1.980 ± 0.685 |
| 677 Q | 0.909 | 4.483 | 0.802 | 2.017 ± 0.659 |
| 688 A | 0.985* | 4.835 | 0.886 | 2.145 ± 0.525 |
| 1263 P | 0.880 | 4.348 | 0.787 | 1.994 ± 0.675 |

M2 selection model: Naive Empirical Bayes (NEB) analysis. Positively selected sites (*: P > 99% ; *: P > 99%). Bayes Empirical Bayes (BEB).

Sites identified under negative selection at p ≤ 0.1. Grey rows: significant sites also identified by SLAC.

MAFFT web server (Katoh and Standley, 2013) using default options and 1PAM = k/2 scoring matrix. Single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs) were assessed by using snippy variant calling pipeline v4.6.0 (https://github.com/teemansnippy). The mapping of the aligned sequences to the reference genome (NC_045512.2) features was created by the software GENEIOUS 2021.0.3 (https://www.geneious.com). Histogram of mutations was generated using a modified code from Lu et al., 2020 (https://github.com/aduplessis/SARS-CoV-2_Guangdong_genomic_epidemiology/).

4.3. Phylogenetic analysis

The previously aligned genomic sequences were used as input for the phylogenetic analysis. The reference genome NC_045512.2 was added as an outgroup and 16 other sequences from B.1.1.28 and B.1.1.33 lineages were added to analyze the phylogeny patterns with the E484K-containing genomes. The inference of the best evolutionary model was performed by ModelTest-NG (Darriba et al., 2020), which identified GTR + I (Generalised time-reversible model + proportion of invariant sites). The tree was built by Bayesian inference in MrBayes v3.2.7 (Huelsenbeck and Ronquist, 2001) with the GTR + I model and 5 million generations. The data convergence was evaluated with Trace v1.7.1 (Rambaut et al., 2018) and tree visualization was created by FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

4.4. Selection analysis

In total, 780 SARS-CoV-2 genomes, collected from Brazil samples between February 1st and December 31st, 2020, available on GISAID, were used for this analysis. After an initial filtering step of undefined nucleotides in the region of spike protein, with deletion of genomes with a N ratio > 0.01, 589 genomes were selected. The second step of sub-sampling, with Augur toolkit (Huddleston et al., 2021) kept 498 genomes (approximately 119 representative genomes per lineage) for the multiple sequence alignment (Table S3). The genomic multiple sequence alignment was performed with MAFFT (Katoh and Standley, 2013) with default options and 1PAM = k/2 scoring matrix. The selection of the genome location for the spike protein was performed with the software UGENE (Okonechnikov et al., 2012), using the ORF coordinates of the NC_045512 reference genome as parameter. GTR + I was inferred by ModelTest-NG (Darriba et al., 2020) as the best evolutionary model for the spike sequences. ModelTest-NG also indicated identical sequences, which were removed resulting in 161 sequences. The Bayesian...
Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We thank the administrators and curators of the GIASA database and research groups across the globe for supporting the rapid and transparent sharing of genomic data during the COVID-19 pandemic. We also thank the Mayor’s Office, Health Department and São Camilo Hospital (Esteio, RS, Brazil), Leonardo Duarte Pascoal and Ana Regina Boll for their work in combating COVID-19 and for supporting the work developed by our research group. A full table acknowledging the authors and corresponding labs submitting sequencing data used in this study can be found in Supplementary Files 1 and 3.

References

Abu-Raddad, L.J., Chemailhity, H., Butt, A.A., 2021. Effectiveness of the BNT162b2 Covid-19 vaccine against the B.1.1.7 and B.1.351 variants. New Engl. J. Med. https://doi.org/10.1056/NEJMoa2104974.

Andreason, E., Piccini, G., Licastro, D., Casalino, L., Johnson, N.V., Paciello, I., Monego, S.D., Pantano, E., Manganaro, N., Manenti, A., Manna, R., Casa, E., Hyseni, L., Benincasa, L., Montomoli, E., Amarou, R.E., Mcellan, J.S., Rappuoli, R., 2020. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. medRxiv, 2020.12.28.20244551.

Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lana, K., Negrón, N., Li, M., Wei, Y., Arwal, G.S., Murphy, A.J., Stahl, N., Yanceopoulos, G.D., Zyatsous, C.A., 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science 369, 1014-1018.

Benvenuto, D., Angeletti, S., Giovanetti, M., Bianchi, M., Pascarella, S., Cauda, R., Ciccozzi, M., Cassone, A., 2020. Evolutionary analysis of SARS-CoV-2: how mutation of non-structural protein 6 (NSP6) could affect viral a tropency. J. Infec. Secur. 81, c24-e7.

Chi, X., Yan, R., Zhang, Jun, Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y., Chen, Z., Guo, Y., Zhang, J., Long, Li, Song, Y., Xie, X., Ren, F., Hu, L., Fan, L., Xu, J., Yu, C., Li, J., Zhou, Q., Chen, W., 2020. A neutralizing human antibody binds to the N-terminal domain of the spike protein of SARS-CoV-2. Science 369, 650-655.

Darriga, D., Pasuda, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2021. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. 37, 291-294.

van Dorp, L., Acman, M., Richard, D., Shaw, L.P., Ford, C.E., Ormond, L., Owen, C.J., Pang, J., Tan, C., Ishihara, F., Ortiz, A.T., Balloux, F., 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infe. Genet. Evol. 40, 104351.

Faria, N., Claro, L.M., Candido, D., Franco, L.A.M., Andrade, P.S., Coletti, T.M., Silva, C.A.M., Prata, N., Carvakhina, G., Pom, L.E., Filho, S., Siqueira, A., Lopes, P.H., Ogah, S.O., Sabino, E.C., 2021. Emergence of Characterization of an Emergent SARS-CoV-2 Lineage in Manaus: Preliminary Findings [WWW Document]. Virological. URL: https://virological.org/t/emergent-characterization-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (accessed 1.14.21).

Grenaney, A.J., Sarr, T.N., Gilchuk, P., Zost, S.J., Binhstein, E., Eoes, A.N., Hilson, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., Dinges, A.S., Nargi, R.S., Sutton, R.E., Suryadevara, N., Rohflad, P.W., Liu, Z., Whelan, S.P.J., Carnahan, H.L., Crowe, J.E., Bloom, J.D., 2021. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host Microbe 29 (1), 44-57.e9.

Gu, H., Chen, Q., Yang, G., He, L., Fan, H., Deng, Y.-Q., Wang, Y., Teng, Y., Zhao, Z., Cui, Y., Li, Yuchang, Li, X. P., Li, J., Zhong, N.-N., Yang, Xiaolan, Chen, S., Guo, Y., Zhao, G., Wang, X., Luo, D.-Y., Wang, H., Yang, Xiao, Li, Yan, G., He, Y., Zhou, X., Geng, S., Sheng, X., Jiang, S., Sun, S., Qin, C.-F., Zhou, Y., 2020. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. Science 369, 1603-1607.

Gupta, R.K., 2021. Will SARS-CoV-2 variants of concern affect the promise of vaccines? Nat. Rev. Immunol. https://doi.org/10.1038/s41577-021-00556-5.

Hodcroft, E.B., Domman, D.B., Oguntuyo, K., Snyder, D.J., Van Diest, M., Demnare, K.H., Schwalm, K.C., Fennig, J., Carroll, J.L., Scott, R.S., Whyte, M.M., Edwards, M.D., Hall, N.C., Kevl, C.G., Vanchiere, J.A., Lee, B., Dwiraddi, D.L., Cooper, S.P., Kamil, J.P., 2021. Emergence in late 2020 of multiple lineages of SARS-CoV-2 spike protein variants affecting amino acid position 677, medRxiv, 2021.02.12.21251658.

Huddleston, J., Hadfield, J., Sibley, T.K., Lee, J., Fay, K., Beisits, M., Harkins, E., Bedford, T., Neher, R.A., Hodcroft, E.B., 2021. Augur: a bioinformatics tool for phylogenetic analyses of human pathogens. J. Open Source Softw. 6, 2906.

Huelenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.
Infection, Genetics and Evolution 93 (2021) 104941

PAG. Ferrareze et al.

Rambaut, A., Holmes, E.C., O’Riordan, L., Pybus, O.G., Barrell, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D.L., Volz, E., 2020b. Preliminary Genomic Characterization of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations [WWW Document]. Virological. URL. https://virological.org/t/preliminary-genomic-characterisation-of-a-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations-563 (accessed 5.17.21).

Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., Appel, C., Rodi-Guinea, L., Roser, M., 2021. Brazil: Coronavirus Pandemic Country Profile [WWW Document]. Our World in Data. URL. https://ourworldindata.org/coronavirus/country/brazil (accessed 5.17.21).

Shu, Y., McCayen, J., 2017. GISAID: global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 22.

Tennissen, J.A., 2008. Positive selection drives a correlation between non-synonymous/ synonymous divergence and functional divergence. Bioinformatics 24 (12), 1421–1425.

Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., Kiyanisti, K., 2020. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. J. Genet. 66, 1075–1082.

Volch, C.M., Francisco, R.d.s, a, Almeida, L.G.P., Cardoso, C.C., Brustolini, O.J., Gerber, A.L., Guimaraes, A.P.d.e, z, Mariano, D., Costa, R.M.d.a, Ferreira, O.C., Covid19-UFRJ Workgroup, INGC Workgroup, Cavalcanti, A.C., Frauchs, T., de Mello, C.M.B., Letia, L.d, C, Gallier, R.M., Paffe, D.S., Castreinations, T.M.P.F., de Vasconcelos, A.T.R., 2021. Genomic characterization of a novel sars-cov-2 lineage from Rio de Janeiro, Brazil. J. Virol. 95 (10).

Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., Toole, S., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., Jackson, B., Nascimento, F.F., Newbold, P., 2020. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 184 (1), 64–75.e11.

Weisblum, Y., Schmidt, F., da Silva, J., Poston, D., Lorenzi, J.C., Mueckf , F., Rutkowska, M., Hoffmann, H.-H., Michalidis, E., Geier, C., Agudelo, M., Cho, A., Wang, J., Ganzymen, T., Zhang, O., Barclay, W., Barrett, J., Volz, E., Bloom, J.D., 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain interactions from mammalian sequence data. Nat. Rev. Genet. 16, 224–236.

World Bank. (2021). Data portals. World Development Indicators. https://databank.worldbank.org/source/world-development-indicators.

World Health Organization. (2021). Coronavirus Disease 2019 (COVID-19) Situation Report – 182. URL. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20210517-covid-19-sitrep-182.pdf?sfvrsn=8.

Worldometer. (2021). Brazil’s COVID-19 Cases and Deaths: A Coronavirus Tracker Source. URL. https://www.worldometers.info/coronavirus/country/brazil/

Worldometers. (2021). Brazil’s COVID-19 Cases and Deaths: A Coronavirus Tracker Source. URL. https://www.worldometers.info/coronavirus/country/brazil/