



Research Article  
Human and Medical Genetics

## Evolutionary history of the SARS-CoV-2 Gamma variant of concern (P.1): a perfect storm

Yuri Yépez<sup>1\*</sup> , Mariana Marcano-Ruiz<sup>1\*</sup> , Rafael S Bezerra<sup>2\*</sup>, Bibiana Fam<sup>1</sup>, João PB Ximenez<sup>2</sup>, Wilson A Silva-Jr<sup>2,3</sup>  and Maria Cátira Bortolini<sup>1</sup> 

<sup>1</sup>Universidade Federal do Rio Grande do Sul, Departamento de Genética, Laboratório de Evolução Humana e Molecular, Porto Alegre, RS, Brazil.

<sup>2</sup>Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, Departamento de Genética, Ribeirão Preto, SP, Brazil.

<sup>3</sup>Instituto de Pesquisa do Câncer de Guarapuava, Guarapuava, PR, Brazil.

### Abstract

Our goal was to describe in more detail the evolutionary history of Gamma and two derived lineages (P.1.1 and P.1.2), which are part of the arms race that SARS-CoV-2 wages with its host. A total of 4,977 sequences of the Gamma strain of SARS-CoV-2 from Brazil were analyzed. We detected 194 sites under positive selection in 12 genes/ORFs: *Spike*, *N*, *M*, *E*, *ORF1a*, *ORF1b*, *ORF3*, *ORF6*, *ORF7a*, *ORF7b*, *ORF8*, and *ORF10*. Some diagnostic sites for Gamma lacked a signature of positive selection in our study, but these were not fixed, apparently escaping the action of purifying selection. Our network analyses revealed branches leading to expanding haplotypes with sites under selection only detected when P.1.1 and P.1.2 were considered. The P.1.2 exclusive haplotype H\_5 originated from a non-synonymous mutational step (H3509Y) in H\_1 of *ORF1a*. The selected allele, 3509Y, represents an adaptive novelty involving *ORF1a* of P.1. Finally, we discuss how phenomena such as epistasis and antagonistic pleiotropy could limit the emergence of new alleles (and combinations thereof) in SARS-CoV-2 lineages, maintaining infectivity in humans, while providing rapid response capabilities to face the arms race triggered by host immuneresponses.

**Keywords:** Gamma, P.1, evolution, COVID-19.

Received: September 30, 2021; Accepted: December 29, 2021.

### Introduction

According to the World Health Organization (WHO) and multiple researchers, the estimated average mortality rate, considering detectable/reported cases, for COVID-19 is lower (2.72%) than the disease caused by MERS-CoV (34.4%) and SARS-CoV (9.6%) (Xiao *et al.*, 2020; ECDC, 2021a,b; Krishnamoorthy *et al.*, 2021; Awadasseid *et al.*, 2021). This number remains low even considering that the number of deaths caused by COVID-19 may be underestimated by 50%, as seen in Wuhan, China (Liu, J. *et al.*, 2021). Despite this relatively low mortality rate, SARS-CoV-2 infection has led to the deaths of 4,219,578 people (WHO, 2021a). Over the past 18 months since the first reported COVID-19 case, the WHO still recognizes that the global public health risks associated with COVID-19 remain very high (WHO, 2021b). Comparatively, SARS-CoV and MERS-CoV infected 8,098 and 2,566 people and killed 774 and 866 people, respectively (WHO, 2003; Alfaraj *et al.*, 2019; WHO, 2020; Petersen *et al.*, 2020).

SARS-CoV, MERS-CoV, and SARS-CoV-2 have high mutation rates ( $0.80\text{--}2.38 \times 10^{-3}$  substitutions per site per year (Zhao *et al.*, 2004; Cotten *et al.*, 2014; Li R *et al.*, 2020). These mutation rates are of the same order of magnitude as other RNA viruses, and can lead to the acquisition of enhanced virulence

and high evolvability, favoring changes in the host and rapid dispersion. A successful zoonotic spillover also depends on the vulnerability of the new host's defenses, and ecological and climatic conditions. Human populations also have cultural habits, with some of them facilitating the transmission of pathogens, *i.e.*, hugs, kisses, sharing food (Olival *et al.*, 2017; Duffy, 2018). Thus, *Homo sapiens* has become a potentially easy target of new pathogens in modern times because of its large population size, urbanization, ease of mobility of people between cities, countries, and continents, and close contact with wild, semi-wild, and domesticated animals. These conditions were in place for SARS-CoV, MERS-CoV, and SARS-CoV-2 so that the interspecific barriers were overcome, and related diseases have been reported (Kan *et al.*, 2005; Zaki *et al.*, 2012; Hedman *et al.*, 2021).

However, there is a notable difference in the outbreak trajectories associated with these  $\beta$ -COVs, specializing in infecting humans and causing severe respiratory syndrome symptoms, as mentioned above. No complete scenario explaining such differences is well understood. However, it is possible to suggest that certain potential drivers, shaped by microevolutionary phenomena, can turn a local epidemic into a global pandemic, as found with SARS-CoV-2: stronger tropism involving host cells, high transmissibility, elevated transmission rates from asymptomatic individuals, substantial viral load, and relatively low lethality all powerful triggers for the emergence of evolutionarily successful viral lineages. All these conditions/factors together represent a "perfect storm".

Send correspondence to Maria Cátira Bortolini. Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Departamento de Genética. Av. Bento Gonçalves, 9500. Prédio 43323, Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br.

\*These authors contributed equally.

COVs may have originated millions of years ago (Wertheim *et al.*, 2013); as such, their natural hosts, species of vertebrates, have been under attack for an equivalent time. The success of zoonotic spillover, that is, transmitting a pathogen from a vertebrate animal to a human and *vice versa*, is a systematic process. Thus, COVs and vertebrates, including primates, are expected to have a long evolutionary history of biological arms races (Meyerson and Sawyer, 2011; Enard *et al.*, 2016; Wang *et al.*, 2020). That is, the host defense system wins at a given evolutionary moment and context; in another, the pathogen attack system wins. There is constant selective pressure on the losing side to change, resulting in these two antagonistic stages alternating perpetually, unless one species involved goes extinct, putting an end to the ongoing arms race. This phenomenon can be recognized as a form of ‘Red Queen’ dynamics (Van Valen, 1973).

Regarding SARS-CoV-2, the coronavirus RaTG13 of the brown bat (*Rhinolophus affinis*) is a potential ancestor (Ji *et al.*, 2020; Li *et al.*, 2020; Wu *et al.*, 2020; Zhou *et al.*, 2020). Their genomes have 97.41% identity (Malaiyan *et al.*, 2021); however, at least five amino acid (aa) substitutions (F486L, Q493Y, S494R, N501D, and Y505H) at critical sites of the Spike (S) glycoprotein receptor-binding domain (RBD) of RaTG13 are crucial for the Wuhan-SARS-CoV-2 lineage to acquire high tropism with its cognate human cell receptor (cell-surface peptidase angiotensin-converting enzyme 2, ACE2; Wan *et al.*, 2020). The first strain of SARS-CoV-2 identified in Wuhan, China, in December 2019, was also characterized as having five critical amino acid differences in its RBD when compared with SARS-CoV (L455Y, F486L, Q493N, S494D, N501T; SARS-CoV-2 and SARS-CoV aa respectively; Wan *et al.*, 2020; Andersen *et al.*, 2020). However, at present, the intermediate host of SARS-CoV-2 remains unknown, and even the origin of SARS-CoV-2 is still controversial; many theories have been propagated, although unfortunately for non-scientific reasons. Despite this, similar to other novel viral pathogens that have caused epidemics or pandemics involving human populations, the overwhelming conclusion is that SARS-CoV-2 can be found in human hosts through a series of unhappy accidental encounters with animals (Rasmussen, 2021). For a more recent review on this topic, see Lytras *et al.* (2021) and Holmes *et al.* (2021).

The S glycoprotein of SARS-CoV-2 contains a cleavage motif for furin proteases (Coutard *et al.*, 2020; Lau *et al.*, 2020), that mediates efficient fusion of the virus with human cell membranes (Yan *et al.*, 2020). In contrast, in SARS-CoV, furin-mediated cleavage of the S glycoprotein does not occur naturally (Simmons *et al.*, 2004). Follis *et al.* (2006) introduced a functional furin cleavage site in the SARS-CoV S glycoprotein and observed potentiation of membrane fusion activity. Lau *et al.* (2020) showed that *in vitro*, SARS-CoV-2 replication is compromised in cells with deletions or point mutations involving the S1/S2 junction. The authors also highlighted the role of natural selection in shaping viral molecular characteristics (Lau *et al.*, 2020).

In April 2020, Fam *et al.* (2020) reported the conservation of 30 ACE2 sites with records relevant for interactions with SARS-CoV-like S glycoproteins in *H. sapiens* populations (Fam *et al.*, 2020). The authors concluded that SARS-CoV-2 has a similar potential to infect humans. Subsequent

investigations corroborated these findings; although ACE2 single polymorphisms have been identified in a number of human populations, none of these SNPs markedly altered interactions between ACE2 and the SARS-CoV-2 S glycoprotein (Hashizume *et al.*, 2021). The contrast between low ACE2 diversity in the target species and the relative diversity between species has been interpreted as a sign of the long evolutionary arms race between CoVs and potential mammalian hosts (Fam *et al.*, 2020). However, it was predicted that successive SARS-CoV-2 mutations would occur, impacting the course of the COVID-19 pandemic, including the dynamics of the evolutionary arms race between the virus and its more recent host, *H. sapiens* (Fam *et al.*, 2020).

The combined impact of these SARS-CoV-2 and human ACE2 characteristics, shaped by evolutionary forces is responsible for the magnitude of the COVID-19 pandemic. For instance, SARS-CoV-2 has high tropism with human ACE2 (Andersen *et al.*, 2020; Wan *et al.*, 2020) and higher transmissibility ( $R_0 = 2.5$ ) than SARS-CoV (2.4), and even higher than the influenza virus (2.0), which caused the 1918 pandemic (Petersen *et al.*, 2020). In addition, 59% of all SARS-CoV-2 transmissions came from asymptomatic individuals (35% from pre-symptomatic individuals and 24% from individuals who never developed symptoms), another powerful trigger for accelerated transmission (Johansson *et al.*, 2021).

In April 2021, the WHO began to recognize the existence of SARS-CoV-2 VOCs and other variants of interest (VOIs) that pose enormous challenges to the management of the outbreak itself, vaccination, and the treatment of symptoms and outcomes linked to the major COVID-19 second/third wave that has plagued several countries more recently (Wang *et al.*, 2020; Awadasseid *et al.*, 2021; WHO and other health agencies).

The first VOC identified was Beta (B.1.351; Nextstrain clade 20H/V2) in South Africa, which appeared in May–August 2020 (Tegally *et al.*, 2020; World Health Organization, 2021c). This derived lineage is characterized by nine changes in the S glycoprotein beyond the 614G allele already present in its parental lineage, which was dominant in South Africa (Tegally *et al.*, 2020).

The WHO and other reports identified several key SARS-CoV-2 S mutations in Beta (B.1.351): D80A, D215G, 241/243del, K417N, E484K, N501Y, D614G, A701V (WHO, 2021c,d,e,f).

Lineage Alpha (B.1.1.7; Nextstrain clade 20I/V1) emerged in southeast England in September–November 2020 and is rapidly spreading toward fixation (Davies *et al.*, 2021a; World Health Organization, 2021b,c). Nine key mutations involving the S glycoprotein have been identified: 69/70del, 144del, N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H (WHO, 2021c,d,e,f).

In early January 2021, Japanese authorities reported that four people from Amazonas, Brazil, had a new SARS-CoV-2 variant. Currently, the WHO has identified the Gamma lineage (B.1.1.28.1, P.1 or Gamma; Nextstrain clade 20J/V3) with the following key S mutations: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, and V1176F. Gamma probably originated in the state of Amazonas in November 2020 and in the capital Manaus. In this city, lineage B.1.1.195 was replaced by B.1.1.28, and

then by B.1.1.28.1 (Gamma) in <2 months (Faria *et al.*, 2021b). Soon after, Gamma quickly spread to other Brazilian states (Faria *et al.*, 2021a, b; WHO, 2021c,d,e,f).

In May 2021, a new VOC was recognized: Delta (B.1.617, Nextstrain clade 21A) which has three sub-lineages: B.1.617.1, B.1.617.2, and B.1.617.3. Key S mutations: L452R, D614G, P681R, ± (E484Q, Q107H, T19R, del157/158, T478K, D950N; WHO, 2021c,d,e,f). Delta lineages were first reported in India in October 2020. The Delta VOC was responsible for the second devastating wave of COVID-19 in India (Ranjan *et al.*, 2021). In June 2021, researchers and WHO members speculated that this VOC could become dominant worldwide because of its increased transmissibility (Campbell *et al.*, 2021; Reuters, 2021).

These VOCs have alleles that confer high transmissibility with some capacity for a second attack (reinfection), even after vaccination (Wibmer *et al.*, 2021; Davies *et al.*, 2021a; Jassat *et al.*, 2021; Davies *et al.*, 2021a; Davies *et al.*, 2021b; Graham *et al.*, 2021; ECDC, 2021a; NERVTAG, 2021; Challen *et al.*, 2021; Bager *et al.*, 2021; Faria *et al.*, 2021a,b; Buss *et al.*, 2021; Taylor, 2021; Martins *et al.*, 2021; Franceschi *et al.*, 2021; Montagutelli *et al.*, 2021; Bogler *et al.*, 2020; Bivins *et al.*, 2020).

In November 2021, the WHO classified the Omicron lineage as a VOC. Omicron is evolving, and the current situation in terms of epidemiology and transmissibility, clinical severity, risk of reinfection and potential impact on diagnostics, vaccines and therapeutics is quite preliminary, but will continue to be refined as more data become widely available (WHO, 2021h).

As genomic screening of SARS-CoV-2 strains expands, it has become possible to identify at least some causes and characteristics that have led to the current COVID-19 pandemic with dramatic consequences for humanity. In addition, several studies have tested natural selection involving the SARS-CoV-2 genome based on the proportion of synonymous to non synonymous substitutions (dN/dS) (Tang *et al.*, 2020; Chaw *et al.*, 2020; Li X. *et al.*, 2020; Sohpal, 2021; Yi *et al.*, 2021). The most prominent signal to emerge from these investigations is the combination of diversifying positive selection and purifying selection, depending on the site considered, within the gene encoding the S glycoprotein. Other studies have compared selective patterns in S with those in other regions of the viral genome. For instance, Koçhan *et al.* (2021) showed that the S gene, unlike the others, had higher dN/dS ratios throughout the evolution of the pandemic.

In the present study, we investigated 4,977 SARS-CoV-2 genomic sequences from Brazil and identified as belonging to the VOC Gamma lineage. Our goal was to determine in greater detail the evolutionary trajectory of this Brazilian autochthonous GammaVOC and its derived sub-lineages, P.1.1, P.1.2.

## Material and Methods

We downloaded 4,977 SARS-CoV-2 genomic sequences from Brazil, already classified as Gamma lineage, and publicly available on the GISAID platform (Elbe and Buckland-Merrett, 2017) (downloaded on May 25, 2021). Next, we aligned the genomic sequences with the Wuhan genome as a reference (NC\_045512.2) using MAFFT v7 software (Katoh *et al.*, 2019) using default parameters.

We chose sequence selection criteria based on lineage assignments made using the PANGO Lineages (PANGOLIN) platform to confirm the sequence lineages (O’Toole *et al.*, 2021). The PANGOLIN tools have robust and reproducible criteria, are constantly revised and updated, and when applied to our dataset, resulted in the conservation of a larger number of sequences and consequently resulted in less loss of variability in the genomic alignment, in contrast to sequence classification criteria based on diagnostic sites or signature mutations (Faria *et al.*, 2021a,b; WHO, 2021c,d,e,f). We also considered possible scenarios in which Gamma diagnostic sites could present selection signals.

Subsequently, we generated two datasets from the initial alignment: (a) only with sequences assigned by PANGOLIN as Gamma VOC (P.1); and (b) with sequences designated as Gamma and also including those designated as the derived lineages P.1.1 and P.1.2. The latter lineage has nine recognized mutations in addition to the P.1 lineage-defining ones: *ORF1ab* (synC1150T, synC1912T, D762G, T1820I), *ORF3a* (D155Y, S180F), *M* (synC26954T), *N* (synC28789T), and S glycoprotein (A262S) (de Almeida *et al.*, 2021). P.1.1 has 21 mutations, all found in Gamma. One difference involves two other alleles of the *N* gene, 203K and 204R, found in Gamma but not in the P.1.1 sub-lineage. The dataset with Gamma and derived lineages will be referred to as a “complete dataset” to facilitate reading. Additionally, for phylogeographic analyses, we worked with a third dataset that included more P.1.1 sequences deposited in the GISAID platform after May 25 (sequences downloaded until June 17, 2021).

## Selection tests

We used ModelFinder (Kalyaanamoorthy *et al.*, 2017) to select the best-fit substitution model and inferred maximum likelihood trees of genomic sequence alignments of the two datasets using IQ-TREE2 with default settings (Minh *et al.*, 2013). We then filtered the alignment by gene and open reading frame (ORF) using SARS-CoV-2 sequences available in NCBI gene records. We performed a series of selection analyses using HYPHY 2.5.31 (Kosakovsky *et al.*, 2020). We implemented three site-level methods: FUBAR (Murrell *et al.*, 2013) and FEL (Kosakovsky and Frost, 2005), to detect sites subject to pervasive positive (diversifying) and negative selection, and MEME (Murrell *et al.*, 2012), a method at the codon level to detect pervasive and episodic positive selection (diversifying). The MEME, FUBAR, and FEL tests use the ratio of non-synonymous to synonymous substitutions (dN/dS) as metrics. We established a  $p$ -value  $\leq 0.1$  for FEL and MEME for statistical significance thresholds and a posterior probability threshold  $\geq 0.9$  for FUBAR (Spielman *et al.*, 2019).

Notably, the dN/dS ratio was originally developed for the analysis of genetic sequences of divergent species. However, viral nucleotide substitution rates can be up to millions of times greater than those of their hosts. This rapid evolution is mainly due to high mutation rates. Despite controversies (Kryazhimskiy and Plotkin, 2008), approaches based on the dN/dS ratio are also appropriate for use in the context of virus populations, including to rescue the evolutionary history of SARS-CoV-2 (Kumar *et al.*, 2020).

## Network analysis

Haplotype networks are used to visualize genealogical relationships at the intraspecific level and to make inferences regarding the biogeography and history of SARS-CoV-2 Gamma populations. We used this methodology to visualize potential signs of expansion associated with lineages and/or groups of sequences (haplotypes) that carry positively selected alleles. Specifically, to analyze the networks in conjunction with the selection tests (FUBAR, FEL, and MEME), we used the following criteria: we chose haplotypes composed of >7 sequences, and subsequently, the mutations that led to these haplotypes from the main haplotype(s) were noted.

Haplotype detection was performed using DnaSP v6 software (Rozas *et al.*, 2017), considering only the variable sites and disregarding gaps or missing data. Subsequently, we constructed networks using NETWORK software v10.2 (Fluxus Technology, 2020) using the median-joining algorithm (Bandelt *et al.*, 1995). Maximum parsimony calculations (Polzin and Daneschmand, 2003) were used to identify unnecessary median vectors and links. Additionally, to simplify the complex network, we employed the “Frequency>1” criterion (Fluxus Technology, 2020), which ignores unique sequences in the dataset.

## Phylogeography of P.1.1 and P.1.2

To better understand the spread of P.1.1 and P.1.2 variants across spatiotemporal scales within Brazil, we employed a discrete diffusion model (Lemey *et al.*, 2009) that maps phylogenetic nodes to their inferred locations of origin, as implemented in the software package BEAST v1.8.4. (Drummond *et al.*, 2012). We worked with two separate data sets, one for P.1.1 (collected until June 17, 2021) and another for P.1.2 (collected until June 6, 2021), both available on the GISAID platform. More P.1.1 sequences were identified, but no additional P.1.2 sequences considering those already downloaded on May 25, 2021. The genomes were classified into lineages using PANGOLIN, as previously described. The sequences classified as P.1.1 and P.1.2 were selected for a total of 18 and 224 sequences in each dataset, respectively. Subsequently, alignments were performed for each lineage using MAFT v7 as described above.

We employed the GTR + G model of nucleotide substitution, a uniform strict molecular clock ( $8-10 \times 10^{-4}$  substitutions/site/year) (Naveca *et al.*, 2021), and assumed an exponential population size growth coalescent process. We considered one discretization scheme considering the state’s capital where the samples were taken, using an asymmetric substitution model allowing Bayesian stochastic search variable selection (BSSVS). Markov chain Monte Carlo (MCMC) was run for >80 million generations and sampled every 10,000 generations. The convergence and mixing properties were inspected using Tracer 1.7.1 (Rambaut *et al.*, 2018). After discarding 10% of sampled trees as burn-in, a maximum clade credibility (MCC) tree was obtained using TreeAnnotator 1.8.4 included in the BEAST v1.8.4 package. We used the resulting MCC tree as input for Spread3 v0.9.7.1 (Spatial Phylogenetic Reconstruction of Evolutionary Dynamics using Data-Driven Documents, Bielejec *et al.*, 2018), which allowed us to analyze and visualize the reconstructions resulting from Bayesian inference of the variants between February and June 2021.

## Results

We detected certain changes involving the sites that define the Gamma VOC according to Faria *et al.* (2021a, b) and the WHO (2021c) (Table 1). For instance, of the 4,977 sequences reported as Gamma in GISAID, we observed that 1,442 did not have a complete set of key mutations (alleles) that characterize it: ~29% relative to the initial number of sequences. Notably, 922 (18.5%) sequences presented a reverse mutation at position 417, that is, lysine residues (K) rather than threonines (T) in the S glycoprotein. We also observed 37 (~0.68%) sequences with the 614D allele (aspartic acid residue present in the Wuhan-SARS-CoV-2 sequence). 417T and 614G are considered diagnostic alleles for Gamma (WHO, 2021b,c,e).

When working with public datasets from various research and diagnostic centers, it is difficult to know whether the loss of one or more diagnostic alleles represents a methodological artifact or a real phenomenon involving the natural evolution of the Gamma lineage. Therefore, the situation is likely to be a combination of both. However, to minimize the chance of error and considering our approach to identify evolutionary footprints, we chose to classify lineages using the PANGOLIN platform (Table 2). The PANGOLIN tool preserves intra-clade diversity, as >93% of the available sequences were classified as belonging to the Gamma lineage. In addition, 6.4% of them could be assigned to other lineages, including lineages derived from Gamma, such as P.1.1 and P.1.2 (Table 2).

Notably, using the PANGOLIN tool, 33 and 902 sequences with alleles 614D and 417K, respectively, were preserved, indicating that these allelic reversions to an ancestral state may represent an actual phenomenon within Gamma.

Table 3 and Table S1 present a comparison of the positively selected sites, considering the two datasets (Gamma and the complete dataset). The same genes in both datasets presented sites under positive selection (*S*, *N*, *M*, *E*, *ORF1a*, *ORF1b*, *ORF3*, *ORF6*, *ORF7a*, *ORF7b*, *ORF8*, and *ORF10*; Table S1).

Using the MEME method, Faria *et al.* (2021b) found 18 sites with positive selection signatures in *N*, *ORF1a*, *ORF1b*, and *ORF3*, including some Gamma S diagnostic sites. Notably, in our analysis, a number of these sites (*e.g.*, E484K) did not exhibit evidence of positive selection.

Interestingly, diagnostic sites that lost the positive selection signal in our study were not entirely fixed, and apparently they were not on the way to becoming fixed within the Gamma clade. This was already noticeable through the data presented in Table 1, but these suggestions become more robust because the diagnostic sites that lose positive selection signatures do not appear to be subject to the powerful action of purifying selection (Table S2).

Our next step was to generate networks involving all 12 genes that showed sites whose diversity seemed to be the result of positive selection. We chose to show those that met the criteria of haplotypes composed of >7 sequences and with some signal of expansion based on network topology for more detail, see Material and Methods). For instance, star-like clusters of nodes surrounding a founder node are classic scenarios that reveal expansion events. It is also known that star phylogeny networks present many rare haplotypes, with one or a few mutational steps from a central haplotype at high frequency (Bandelt *et al.*, 1995). However, because of the frequent transmission of viruses, rapid lineage expansion

**Table 1** – Observed sequences without P.1 (Gamma) mutation signatures\*.

Gene	Gamma signature mutations	Number of sequences without signature mutations	%
<i>ORF1ab</i>	synT733C	5	0.10
	synC2749T	88	1.77
	S1188L	94	1.89
	K1795Q	1	0.02
	del11288-11296(3675-3677) SGF	192	3.86
	synC12778T	116	2.33
	synC13860T	87	1.75
	E5665D	90	1.81
<i>Spike</i>	L18F	6	0.12
	T20N	88	1.77
	P26S	6	0.12
	D138Y	209	4.20
	R190S	38	0.76
	K417T	922	18.53
	E484K	15	0.30
	N501Y	11	0.22
	D614G	37	0.68
	H655Y	4	0.08
T1027I	4	0.08	
<i>N</i>	P80R	48	0.96
	ins28269-28273	-	-
<i>ORF8</i>	E92K	94	1.89

\*According to Faria *et al.* (2021a, b) and WHO, 2021c,d,e,f. Total sequences: 4,977.

**Table 2** – Classification of sequences according to PANGOLIN.

Lineage	Number of sequences	%
B.1	3	0.06
B.1.1	58	1.17
B.1.1.28	92	1.85
B.1.1.74	1	0.02
B.1.560	1	0.02
B.1.566	1	0.02
P.1	4,660	93.63
P.1.1	15	0.30
P.1.2	146	2.93

**Table 3** – Comparison between positively selected sites detected in VOC P.1 (Gamma) and its potential derived-lineages P.1.1 and P.1.2.

Positive selection	Gamma	Gamma, P.1.1, P.1.2
Sites detected by at least one of the methods	197	214
Sites detected by all three methods	77	52
MEME detected sites for at least 1 branch	126	92
Sites detected only by MEME (possible episodic selection)	23	20
Diagnostic sites that are under selection	9	12

involving rare and random alleles can occur due to founder effects (Croucher and Didelot, 2015; Campbell *et al.*, 2021). However, despite the expected scenario involving random phenomena, we hypothesized that the adaptive advantage led, at least in part, to the expansion of lineages within Gamma.

Table 4 shows the sites that met these criteria in both datasets [Gamma and Gamma/P1.1/P.1.2; *S* (L5F, T572I, A845S), *ORF3* (L83F, K16N, L85F, D27Y), *ORF1a* (G150S, G519S, P2046T, L642F, K3353R), *ORF1b* (L314P, A1643V, D1264E, T1774I)], while others were unique to one or another dataset, suggesting potentially different evolutionary pathways from the ancestor (Gamma) and its two more recently derived lineages.

The simplified networks (Figure S1 A-D), considering our complete dataset, revealed a known star-like pattern, as well as some thought-provoking findings for gene/ORFs *S*, *ORF3*, *ORF1a*, and *ORF1b*. Some rare reticulations may be due to parallel mutations, homoplasy, recombination, and/or methodological errors. For example, the synonymous substitution C1818T in the RNA sequence is recurrent in *S* and causes several reticulations (Figure S1 A), but we do not know if it represents a natural SARS-CoV-2 genomic mutational hotspot, or a simple annotation or sequencing error.

Haplotypes 5 and 23 (H\_5 and H\_23, Figure S1 A) have the selected sites A845S and L5F in the *S*. H\_23 is present in all Brazilian regions, whereas H\_5 is present only in the southeast, the most densely populated region of the country. It is noteworthy that the positive selection signal was not lost when considering the derived lineages (Table 4), similar to what happens with the *ORF3* selected sites L83F and D27Y; *ORF1a* selected sites G150S, G519S, P2046T/L, and K3353R, as well as *ORF1b* selected sites L314P, D1264E, and T1774I (Table 4), all of which were present in haplotypes with some level of expansion (Figure S1 B: H\_2, H\_36; Figure S1 C: H\_64, H\_222, H\_79, H\_89; Figure S1 D: H\_11, H\_60, H\_261, respectively). Other branches that led to haplotypes with signs of expansion had sites under positive selection detected only when the complete data were analyzed (Table 4). For example, H\_5 originated from a non-synonymous mutational step (position 10525 of the RNA sequence/ H3509Y) from H\_1 of *ORF1a* (Figure S1 C). *ORF1b* also presented positively selected sites under these conditions (H\_32, 5484, 5483/ K1828T; H\_65, 652/ P218L; H\_117, 246/ K82N).

Notably, some critical diagnostic sites, whose positive selection signals were detected by us (Table S1), did not seem to be relevant in the networks, indicating no apparent sign of expansion (according to our criteria), at least so far (*e.g.*, *ORF1a* T1820I and *ORF3* D155Y diagnostic sites of P.1.2; de Almeida *et al.* 2021).

On the other hand, Table S3 shows that only the P.1.2-derived lineage assembles a more exclusive set of sequences: 27% and 100% of the *ORF3* and *ORF1a* sequences, identified by PANGOLIN as P.1.2, clustered in H\_2 (Figure S1 B) and H\_5 (Figure S1 C). Notably, H\_5 (*ORF1a*) was composed of sequences with the potentially positively selected allele 3509Y. This finding indicated that P.1.2 has at least one site, not previously identified, that characterizes it as a target of natural selection and can endow a favorable fitness to this emerging lineage.

Concomitantly, our analysis revealed a large number of sites under purifying selection (Table S2), which maintain the general SARS-CoV-2 “status quo,” as a specialist to infect humans (Fam *et al.*, 2020).

Phylogeographic analysis of P.1.1 and P.1.2 sequences (Figure S2 A and B) revealed that the former seems to have originated in the state of Goiás (central-west region, CW) and spread mainly to the southern (S) and south-eastern (SE) states, not reaching other areas; however, the number of sequences used in the analysis of P.1.1 was low, and this result should be taken with caution when making larger inferences. Variant P.1.2 seems to have originated in Rio Grande do Sul, the southernmost state in Brazil, with a broader distribution, although mostly restricted to the country’s east. Notably, H\_5 (*ORF1a*), composed only of P.1.2 3509Y sequences, was present in the southeast and northeast regions of Brazil (Figure S1 C). This result suggested that H3509Y of *ORF1a* occurred after P.1.2 had dispersed from the S to the southwest (SW) and northeast (NE) regions.

## Discussion

The perfect storm described herein, culminating in the ongoing tragedy of the COVID-19 pandemic, was only possible because of particular evolutionary events in the trajectory of SARS-CoV-2 that maintained its high transmissibility and relatively low lethality. The dominance of certain SARS-CoV-2 lineages over time relative to others is robust evidence of this fact and that the evolutionary arms race between SARS-CoV-2 and its host is in full swing. In this scenario, we must also compute the “*Homo sapiens* reaction,” including natural immune responses (probably shaped for millions of years due to likely recurrent attacks of CoVs; Meyerson and Sawyer, 2011; Enard *et al.*, 2016; Wang W. *et al.*, 2020) and those induced by current large-scale vaccine deployment, among other containment and pharmacological measures. In addition, Gamma was found in Brazil, a territory already swept by other SARS-CoV-2 lineages; the interaction between different sets of Gamma  $\times$  non-Gamma lineages could also play an important role in determining whether a Gamma lineage would be able to expand.

Because of limited genome size, as seen in viruses, relatively few nucleotide sites are free to vary, despite high mutation rates (Eigen, 1996; Holmes, 2003). Thus, convergent evolution (*e.g.*, the parallel occurrence of identical alleles in distant lineages) is relatively common among RNA viruses (Cuevas *et al.*, 2002). Weinreich *et al.* (2006) demonstrated how many mutational pathways are repeatedly selected in pathogenic microorganisms. The phenomena of epistasis (non-additive interactions among alleles) also dictates the complex course of viral evolution (Holmes and Rambaut, 2004; Fragata *et al.*, 2019). In the presence of epistasis, the fitness effect of a combination of multiple non-interacting alleles of the same or related genes/pathways corresponds to the sum of fitness effects of the individual alleles (Ferretti *et al.*, 2018; Østman *et al.*, 2012; Fragata *et al.*, 2019). Thus, epistasis limits the possible routes to high fitness, and orders which sequences of consecutive mutations are feasible, but can also open new functional evolutionary paths, promoting adaptive novelty (Usmanova *et al.*, 2015; Starr and Thornton, 2016;

**Table 4** – Haplotypes with more than seven sequences, a signal of expansion based on networks, and alleles in sites under positive selection.

Gamma sequences				Gamma, P.1.1, and P.1.2 sequences			
Spike							
Haplotype	N	RNA sequence position/ Aminoacid change	Brazilian region	Haplotype <sup>1</sup>	N	RNA sequence position/ Aminoacid change	Brazilian region
6	12	<b>1715 /T572I</b>	SE	5	67	<b>2533/A845S</b>	SE
15	66	<b>2533 /A845S</b>	SE	10	9	80/A27V	S, SE, N, NE
19	30	<b>13/L5F</b>	SE, S, CW, N, NE	13	12	<b>1715/T572I</b>	SE
45	31	1841 /G614D	S	23	31	<b>13/L5F</b>	SE, S, CW, N, NE
				40	9	35/S12F	SE, S
				101	9	2533/A845S	SE
ORF3							
2	161	<b>249/L83F</b>	SE, S, CW	2	226	<b>249/L83F</b>	CW, S, SE
11	12	161/A54V	SE, NE	3	17	<b>48/K16N</b>	SE
13	16	604/V202L	SE	20	10	<b>255/L85F</b>	SE
15	10	<b>255/L85F</b>	SE	36	33	<b>79/D27Y</b>	SE
19	27	97/A33S	SE				
28	15	<b>48/K16N</b>	SE				
32	33	<b>79/D27Y</b>	SE				
100	8	524/T175I	SE				
ORF1a							
26	76	9626/A3209V	SE, NE, S	5	9	10525/H3509Y	SE, NE
35	15	<b>448/G150S</b>	SE, NE	64	15	<b>448/G150S</b>	SE, NE
183	12	12521/T4174I	SE, NE, CW, N	79	30	6137/P2046L <sup>2</sup>	SE, NE, CW
204	10	<b>1555/G519S</b>	SE	89	19	<b>10058/K3353R</b>	SE, NE
206	13	<b>6136/P2046T</b>	SE	222	10	<b>1555/G519S</b>	SE, CW
210	9	<b>1924/L642F</b>	SE, CW	224	13	<b>6136/P2046T</b> <sup>2</sup>	SE
173	17	<b>10058/K3353R</b>	SE, NE	228	9	<b>1924/L642F</b>	SE, CW
ORF1b							
2	50	<b>941/L314P</b>	SE,N,CW,S	11	56	<b>941/L314P</b>	CW, SE, N, S
21	9	<b>4928/A1643V</b>	SE, N, S	32	8	5484/K1828T <sup>4</sup>	SE
43	74	<b>3792/D1264E</b>	S, SE	39	9	<b>4928/A1643V</b>	SE, S, NE
253	8	<b>5321/T1774I</b>	SE	59	8	5483/K1828T <sup>4</sup>	SE
				60	74	<b>3792/D1264E</b>	S, SE, CW, NE
				65	35	653/P218L	SE, N, CW, NE
				117	19	246/K82N	SE
				135	11	<b>5321/T1774I</b> <sup>3</sup>	SE
				261	8	<b>5321/ T1774I</b> <sup>3</sup>	SE

N: Number of sequences present in each Haplotype. Brazilian regions: SE: Southeast; S: South; Central-West: CW; N: North; NE: Northeast. In bold are the changes in common between both datasets. <sup>1</sup>Networks with haplotypes considering the complete dataset (Gamma plus P1.1 and P1.2 sequences) can be seen in Figure S1 A-D. <sup>2</sup>A non-synonymous mutation in different position at the same codon. <sup>3</sup>While one mutation clearly occurs in a sequence of an ancestral haplotype still existing, the other comes from a median vector automatically generated by the software. A median vector represents potential ancestral sequence/haplotype, but not represented in present sampling. <sup>4</sup>A non-synonymous mutation in different position at the same codon.

Ferretti *et al.*, 2016; Fragata *et al.*, 2019). Sanjuán *et al.* (2004) showed details of how the architecture of fitness in RNA viruses depends on epistasis. Recently, epistasis has been cited as a phenomenon behind the rapid and successful dispersion of the Omicron variant (BBC News Mundo, 2021). Another important phenomenon in the evolution of viral genomes is antagonistic pleiotropy, *i.e.*, mutations beneficial in one host may be deleterious in others. Antagonistic pleiotropy may limit the range of adaptations and promote the evolution of specialization (Santiago *et al.*, 2009). This concept is different from that of classic pleiotropy, characterized by genes/ORFs that affect more than one independent trait, a mechanism that promotes substantial modulation of variation (Wagner and Zhang, 2011), which likely also impacts the evolution of SARS-CoV-2. Below, we discuss how examples presented in this work illustrate the action of such mechanisms.

It is known that a tyrosine (Y) residue at position 501 (located in the receptor-binding domain, RBD, of the S protein), rather than an asparagine (N; present in Wuhan SARS-CoV-2 genome sequence), is present in Alpha, Beta, and Gamma VOCs, potentially creating opportunities for these to become more infectious and partially resistant to therapeutics blocking RBD-ACE2 interactions (Liu H. *et al.*, 2021; Narayanan and Procko, 2021). 501Y endows these lineages with the ability to escape from therapeutically relevant antibodies and the host immune system (Wibmer *et al.*, 2021). Other derived lineages also harbor the N501Y mutation, signaling its recurrence (Lemmermann *et al.*, 2021). VOCs Beta and Gamma share an additional RBD mutation, K417N/T (Abdoolland de Oliveira, 2021), while UK researchers detected cases of Alpha lineages with the 484K allele, which also characterizes the Beta and Gamma lineages (Shamsian, 2021; Faria *et al.*, 2021b). Some studies have shown that lineages containing the 417N allele are more effective in escaping antibody neutralization than Gamma, which presents the 417T allele (Garcia-Beltran *et al.*, 2021; Liu Y. *et al.*, 2021). In addition, the 484K allele creates a new binding site for the amino acid at position 75 in human ACE2. This interaction seems stronger than the binding between ACE2 and the original main site located at position 501 (Ferrareze *et al.*, 2021; Nelson *et al.*, 2021). Other studies have shown that 484K reduces neutralization by polyclonal antibodies (Greaney *et al.*, 2021; Oliveira *et al.*, 2021). Oliveira *et al.* (2021) identified an immune dominant epitope (S<sub>415-429</sub>) recognized by 68% of sera from convalescent Brazilians infected with the ancestral SARS-CoV-2 lineage. This immune dominant RBD region harbors a mutational hotspot site at position 417. The same authors also performed simulations that indicated impaired RBD binding to previous infection- or vaccine-induced neutralizing antibodies in both Beta (417N) and Gamma (417T) VOCs (Oliveira *et al.*, 2021). It is noteworthy that 614G started to appear as a diagnostic allele for VOCs Alpha, Beta, and Gamma only in the WHO report of April 27 (2021 g). Many recurring mutations lead to the exchange of amino acids D > G at position 614 of the SARS-CoV-2 S glycoprotein. Lineages with the 614G allele were rare before March 2020, but became dominant after May 2020, suggesting that the 614G allele might improve viral fitness (Plante *et al.*, 2020). More recently, Xu *et al.* (2021) reported that the 614G allele could lower the energy barrier

for conformational transformation from the RDB closed pre-fusion state to the fusion-prone open state, resulting in even greater affinity of SARS-CoV-2 S protein binding to ACE2. The SARS-CoV-2 lineage from Wuhan harbors the D614 allele.

Some positively selected sites have also been highlighted in studies by the GISAID platform group that applied the MEME and FEL methods to a database of approximately 390,000 sequences per gene/ORF (Gamma sequences included). This team of researchers detected 213 sites with positive selection signatures, 27% of which were located in S; the majority had the same aa changes as we observed (Pond, 2020). For example, the positively selected site T175I (*ORF3*; Table S1), also referred to as a site where epitopes overlap, the minor allele (isoleucine-I) is found on other continents (Pond, 2020), indicating that it is recurrent in several lineages in addition to Gamma. The recurrence of identical alleles in lineages which are otherwise relatively phylogenetically distant appears to be a constant.

All of these studies reinforce the functionality of these positions, and the same mutational pathways are repeatedly selected for. We found in Gamma sequences ancestral alleles at all of these sites (417, 484, 501, and 614), but in different combinations. In other words, in most cases, there are alternations between alleles (amino acid residues) present in the Wuhan sequence and derived alleles. Our results also showed no negative selection signals involving these positions (Table S2), indicating a certain margin for the evolvability (evolutionary capacity) of the SARS-CoV-2 S protein, but within an evolutionary trajectory also modulated by the phenomena of epistasis and pleiotropy.

Relatively few alternatives can represent adaptive novelties in S (for example glutamine (Q) at position 484 in Delta). The potentially positively selected allele 3509Y (*ORF1a*) in nine P.1.2 sequences that form an expanding haplotype also illustrates another adaptive novelty (Figure S1 C). The role of CoV *ORF1a* does not seem to be limited to viral transcription and replication. Graham *et al.* (2008) demonstrated its function in virulence, virus–cell interactions, and alterations to virus–host responses. More recent investigations (Emam *et al.*, 2021) demonstrate that proteins encoded by *ORF1a* and *ORF1b* together (*ORF1ab*) are involved in SARS-CoV-2 pathogenicity and infectivity. Only functional studies will define whether there is a gain in fitness in P.1.2 due to 3509Y. These findings suggest a limitation to the emergence of variability in SARS-CoV-2 modulated by positive selection involving critical positions, as well as the likely existence of compensatory alleles due to epistasis and/or pleiotropy (both antagonistic and classical).

We also observed that many lineages with potentially advantageous alleles succumbed, while others, for no obvious or apparent reasons, expanded (Figure S1 A-D). This finding clearly illustrates the effect of stochastic events (*i.e.*, a rapid expansion of lineages with rare neutral alleles occurring due to founder effects after they reach a large urban center), already predicted to be important in viral evolution (Croucher and Didelot, 2015; Campbell *et al.*, 2021). These two powerful forces (natural selection and random events) are not mutually exclusive because a variant under selection can spread more rapidly in urban and densely populated areas.



Finally, our findings should be considered with caution. For instance, although widely used in population studies involving organisms with high mutation rates (Kumar *et al.*, 2020), methods based on the dN/d S ratio at the population level may result in false-positive results (Kryazhimskiy and Plotkin, 2008). In addition, the quantitative and qualitative complexity of positive selection regimes using the present population genomic data cannot be accessed through our analysis.

## Acknowledgments

This research was supported financially by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Finance code 001), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), and the Instituto de Pesquisa do Câncer de Guarapuava. The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

## Conflicts of Interest

The authors declare that there are no conflicts of interest that could be understood as prejudicial to this research.

## Author Contributions

YY, MM-R, and RSB contributed to the conceptualization, investigation, and formal analysis; JPBX and BF contributed to the investigation and formal analysis; WAS-Jr contributed to supervision (lead), conceptualization, investigation, writing-review, and editing; MCB contributed to supervision (lead), conceptualization, investigation, writing-original draft, and writing-review and editing.

## References

- Abdool Karim S and de Oliveira T (2021) New SARS-CoV-2 variants - clinical, public health, and vaccine implications. *N Engl J Med* 384:1866-1868.
- Alfaraj SH, Al-Tawfiq JA, Assiri AY, Alzahrani NA, Alanazi A and Memish ZA (2019) Clinical predictors of mortality of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: A cohort study. *Travel Med Infect Dis* 29:48-50.
- Andersen KG, Rambaut A, Lipkin WI, Holmes EC and Garry RF (2020) The proximal origin of SARS-CoV-2. *Nat Med* 26:450-452.
- Awadasseid A, Wu Y, Tanaka Y and Zhang W (2021) SARS-CoV-2 variants evolved during the early stage of the pandemic and effects of mutations on adaptation in Wuhan populations. *Int J Biol Sci* 17:97-106.
- Bager P, Wohlfahrt J, Fonager J, Albertsen M, YssingMichaelsen T, HoltenMøller C, Ethelberg S, Legarth R, Fischer Button MS, Gubbels SM *et al.* (2021) Increased risk of hospitalization associated with infection with SARS-CoV-2 Lineage B.1.1.7 in Denmark. SSRN Electron J. DOI: 10.2139/ssrn.3792894.
- Bandelt HJ, Forster P, Sykes BC and Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-753.
- Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A and Lemey P (2016) Spread3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol* 33:2167-2169.
- Bivins A, Greaves J, Fischer R, Yinda KC, Ahmed W, Kitajima M, Munster VJ and Bibby K (2020) Persistence of SARS-CoV-2 in water and wastewater. *Environ Sci Technol Lett* 7:937-942.
- Bogler A, Packman A, Furman A, Gross A, Kushmaro A, Ronen A, Dagot C, Hill C, Vaizel-Ohayon D, Morgenroth E *et al.* (2020) Rethinking wastewater risks and monitoring in light of the COVID-19 pandemic. *Nat Sustain* 3:981-990.
- Buss LF, Prete CA, Abraham CM, Mendrone JrA, Salomon T, Almeida-Neto C, de França RFO, Belotti MC, Carvalho MPS, Costa AG *et al.* (2021) Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* 371:288-292.
- Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, Pavlin B, Vandemaele K, Kerkhove MD, Van Jombart T *et al.* (2021) Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Euro Surveill* 26:2100509.
- Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K and Danon L (2021) Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. *BMJ* 372:n579.
- Chaw SM, Tai JH, Chen SL, Hsieh CH, Chang SY, Yeh SH, Yang WS, Chen PJ and Wang HY (2020). The origin and underlying driving forces of the SARS-CoV-2 outbreak. *J Biomed Sci* 27:73.
- Cotten M, Watson SJ, Zumla AI, Makhdoom HQ, Palser AL, Ong SH, Al Rabeeah A, Alhakeem RF, Assiri A, Al-Tawfiq JA *et al.* (2014) Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* 5:e1062-13.
- Coutard B, Valle C, Lamballerie X, Canard B, Seidah NG and Decroly E (2020) The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* 176:104742.
- Croucher NJ and Didelot X (2015) The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbiol* 23:62-67.
- Cuevas JM, Elena SF and Moya A (2002) Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162:533-542.
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD *et al.* (2021a) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372:eabg3055.
- Davies NG, Jarvis CI, CMMID COVID-19 Working Group, Edmunds WJ, Jewell NP, Diaz-Ordaz K and Keogh RH (2021b) Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 593:270-274.
- Drummond AJ, Suchard MA, Xie D and Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969-1973.
- Duffy S (2018) Why are RNA virus mutation rates so damn high? *PLoS Biol* 16:e3000003.
- Eigen M (1996) On the nature of virus quasispecies. *Trends Microbiol* 4:216-218.
- Elbe S and Buckland-Merrett G (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 1:33-46.
- Emam M, Oweda M, Antunes A and El-Hadidi M (2021) Positive selection as a key player for SARS-CoV-2 pathogenicity: Insights into ORF1ab, S and E genes. *Virus Res* 302:198472.
- Enard D, Cai L, Gwennap C and Petrov DA (2016) Viruses are a dominant driver of protein adaptation in mammals. *Elife* 5:e12469.

- Fam BSO, Vargas-Pinilla P, Amorim CEG, Sortica VA and Bortolini MC (2020) ACE2 diversity in placental mammals reveals the evolutionary strategy of SARS-CoV-2. *Genet Mol Biol* 43:e20200104.
- Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT *et al.* (2021b) Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372:815-821.
- Ferrareze PAG, Franceschi VB, Mayer A de M, Caldana GD, Zimmerman RA and Thompson CE (2021) E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *Infect Genet Evol.* 93:104941.
- Ferretti L, Schmiegel B, Weinreich D, Yamauchi A, Kobayashi Y, Tajima F and Achaz G (2016) Measuring epistasis in fitness landscapes: The correlation of fitness effects of mutations. *J Theor Biol* 396:132-143.
- Ferretti L, Weinreich D, Tajima F and Achaz G (2018) Evolutionary constraints in fitness landscapes. *Heredity (Edinb)* 121:466-481.
- Follis KE, York J and Nunberg JH (2006) Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* 350:358-369.
- Fragata I, Blanckaert A, Louro MAD, Liberles DA and Bank C (2019) Evolution in the light of fitness landscape theory. *Trends Ecol Evol* 34:69-82.
- Franceschi VB, Caldana GD, Perin C, Horn A, Peter C, Cybis GB, Ferrareze PAG, Rotta LN, Cadejani FA, Zimmerman RA *et al.* (2021) Predominance of the SARS-CoV-2 lineage P.1 and its sublineage P.1.2 in patients from the metropolitan region of Porto Alegre, Southern Brazil in March 2021: A phylogenomic analysis. medRxiv. DOI: 10.1101/2021.05.18.21257420.
- Garcia-Beltran WF, Lam EC, Denis KSt, Nitido AD, Garcia ZH, Hauser BM, Feldman J, Pavlovic MN, Gregory DJ, Poznansky MC *et al.* (2021) Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 184:2372-2383.e9.
- Graham MS, Sudre CH, May A, Antonelli M, Murray B, Varsavsky T, Kläser K, Canas LS, Molteni E, Modat M *et al.* (2021) Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: An ecological study. *Lancet Public Health* 6:e335-e345.
- Graham RL, Sparks JS, Eckerle LD, Sims AC and Denison MR (2008) SARS coronavirus replicase proteins in pathogenesis. *Virus Res* 133:88-100.
- Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY and Bloom JD (2021) Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29:463-476.e6.
- Hashizume M, Gonzalez G, Ono C, Takashima A and Iwasaki M (2021) Population-specific ACE2 single-nucleotide polymorphisms have limited impact on SARS-CoV-2 Infectivity *In Vitro*. *Viruses* 13:67.
- Hedman HD, Krawczyk E, Helmy YA, Zhang L and Varga C (2021) Host diversity and potential transmission pathways of SARSCoV-2 at the human-animal interface. *Pathogens* 10:180.
- Holmes EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol* 11:543-546.
- Holmes EC and Rambaut A (2004) Viral evolution and the emergence of SARS coronavirus. *Philos Trans R Soc Lond B Biol Sci* 359:1059-1065.
- Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF, Doherty PC *et al.* (2021) The origins of SARS-CoV-2: A critical review. *Cell* 184:4848-4856.
- Jassat W, Mudara C, Ozougwu L, Tempia S, Blumberg L, Davies MA, Pillay Y, Carter T, Morewane R, Wolmarans *et al.* (2021) Increased mortality among individuals hospitalised with COVID-19 during the second wave in South Africa. medRxiv. DOI: 10.1101/2021.03.09.21253184.
- Ji W, Wang W, Zhao X, Zai J and Li X (2020) Cross-species transmission of the newly identified coronavirus 2019-nCoV. *J Med Virol* 92:433-440.
- Johansson MA, Quandelacy TM, Kada S, Prasad PV, Steele M, Brooks JT, Slayton R B, Biggerstaff M and Butler JC (2021) SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Netw Open* 4:e2035057.
- Kalyanamoorthy S, Minh BQ, Wong TKF, von Haeseler A and Jermin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587-589.
- Kan B, Wang M, Jing H, Xu H, Jiang X, Yan M, Liang W, Zheng H, Wan K, Liu Q *et al.* (2005) Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. *J Virol* 79:11892-11900.
- Katoh K, Rozewicki J and Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160-1166.
- Koçhan N, Eskier D, Suner A, Karakulah G and Oktay Y (2021) Different selection dynamics of S and RdRp between SARS-CoV-2 genomes with and without the dominant mutations. *Infect Genet Evol* 91:104796.
- Kosakovsky SLP and Frost SDW (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22:1208-1222.
- Kosakovsky SLP, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A *et al.* (2020) HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol* 37:295-299.
- Krishnamoorthy P, Raj AS, Roy S, Kumar NS and Kumar H (2021) Comparative transcriptome analysis of SARS-CoV, MERS-CoV, and SARS-CoV-2 to identify potential pathways for drug repurposing. *Comput Biol Med* 128:104123.
- Kryazhimskiy S and Plotkin JB (2008) The population genetics of dN/dS. *PLoS Genet* 4:e1000304.
- Kumar R, Verma H, Singhvi N, Sood U, Gupta V, Singh M, Sood U, Kumari R, Hira P, Nagar S *et al.* (2020) Comparative genomic analysis of rapidly evolving SARS-COV-2 reveals mosaic pattern of phylogeographical distribution. *mSystem* 5:e00505-20.
- Lau S-Y, Wang P, Mok BWY, Zhang AJ, Chu H, Lee ACY, Deng S, Chen P, Chan K-H, Song W *et al.* (2020) Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect* 9:837-842.
- Lemey P, Rambaut A, Drummond A J and Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5:e1000520.
- Lemmermann N, Lieb B, Laufs T, Renzaho A, Runkel S, Kohlen W, Linke M, Gerber S, Schweiger S, Michel A *et al.* (2021) SARS-CoV-2 genome surveillance in Mainz, Germany, reveals convergent origin of the N501Y spike mutation in a hospital setting. medRxiv. DOI: 10.1101/2021.02.11.21251324.
- Li R, Qiao S and Zhang G (2020) Analysis of angiotensin-converting enzyme 2 (ACE2) from different species sheds some light on cross-species receptor usage of a novel coronavirus 2019-nCoV. *J Infect* 80:469-496.
- Li X, Giorgi EE, Marichanegowda MH, Foley B, Xiao C, Kong XP, Chen Y, Gnanakaran S, Korber B and Gao F (2020) Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv* 6:eabb9153.
- Liu H, Wei P, Zhang Q, Chen Z, Aviszus K, Downing W, Peterson S, Reynoso L, Downey GP, Frankel SK *et al.* (2021) 501Y.

- V2 and 501Y.V3 variants of SARS-CoV-2 lose binding to bamlanivimab *in vitro*. *MABs* 13:1919285.
- Liu J, Zhang L, Yan Y, Zhou Y, Yin P, Qi J, Wang L, Pan J, You J, Yang J *et al.* (2021) Excess mortality in Wuhan city and other parts of China during the three months of the covid-19 outbreak: Findings from nationwide mortality registries. *BMJ* 372:n415.
- Liu Y, Liu J, Xia H, Zhang X, Fontes-Garfias CR, Swanson KA, Cai H, Sarkar R, Chen W, Cutler M *et al.* (2021) Neutralizing activity of BNT162b2-elicited serum. *N Engl J Med* 384:1466-1468.
- Lytras S, Xia W, Hughes J, Jiang X and Robertson DL (2021) The animal origin of SARS-CoV-2. *Science* 373:968-970.
- Malaiyan J, Arumugam S, Mohan K and Radhakrishnan G (2021) An update on the origin of SARS-CoV-2: Despite closest identity, bat (RaTG13) and pangolin derived coronaviruses varied in the critical binding site and O-linked glycan residues. *J Med Virol* 93:499-505.
- Martins AF, Zavascki AP, Wink PL, Volpato FCZ, Monteiro FL, Rosset C, De-Paris F, Ramos ÁK and Barth AL (2021) Detection of SARS-CoV-2 lineage P.1 in patients from a region with exponentially increasing hospitalisation rate, February 2021, Rio Grande do Sul, Southern Brazil. *Euro Surveill* 26:2100276.
- Meyerson NR and Sawyer SL (2011) Two-stepping through time: mammals and viruses. *Trends Microbiol* 19:286-294.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A and Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530-1534.
- Montagutelli X, Prot M, Levillayer L, Salazar EB, Jouvion G, Conquet L, Donati F, Albert M, Gambaro F, Behillil S *et al.* (2021) The B.1.351 and P.1 variants extend SARS-CoV-2 host range to mice. *bioRxiv*. DOI: 10.1101/2021.03.18.436013.
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL and Scheffler K (2013) FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol* 30:1196-1205.
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K and Pond SLK (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764.
- Narayanan KK and Procko E (2021) Deep mutational scanning of viral glycoproteins and their host receptors. *Front Mol Biosci* 8:636660.
- Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S and Soon-Shiong P (2021) Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escap. *bioRxiv*. DOI: 10.1101/2021.01.13.426558.
- O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B *et al.* (2021) Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* 7:veab064.
- Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL and Daszak P (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646-650.
- Oliveira JR, Machado RRG, Arcuri HA, Magawa JY, Daher IP, Urbanski AH, Schmitz GJH, Silva RCV, Durigon EL, Boscardin SB *et al.* (2021) Immunodominant B cell epitope in a hotspot mutation site and mechanism of immune escape for SARS-CoV-2. *medRxiv*. DOI: 10.1101/2021.03.11.21253399.
- Østman B, Hintze A and Adami C (2012) Impact of epistasis and pleiotropy on evolutionary adaptation. *Proc Biol Sci* 279:247-256.
- Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, Storgaard M, Khalili S Al and Simonsen L (2020) Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis* 20:e238-e244.
- Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR *et al.* (2020) Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592:7852.
- Polzin T and VahdatiDaneshmand S (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett* 31:12-20.
- Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901-904.
- Ranjan R, Sharma A and Verma MK (2021) Characterization of the second wave of COVID-19 in India. *medRxiv*. DOI: 10.1101/2021.04.17.21255665.
- Rasmussen AL (2021) On the origins of SARS-CoV-2. *Nat Med* 27:9.
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE and Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299-3302.
- Sanjuán R, Moya A and Santiago FE (2004) The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc Natl Acad Sci U S A* 101:15376-15379.
- Santiago FE, Patricia AR and Jasna L (2009) The evolution of viruses in multi-host fitness landscapes. *Open Virol J* 3:1-6.
- Shamsian N (2021) One small step for man... *J Wound Care* 30:91-92.
- Simmons G, Reeves JD, Rennekamp AJ, Amberg SM, Piefer AJ and Bates P (2004) Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc Natl Acad Sci U S A* 101:4240-4245.
- Sohpal VK (2021) Comparative study: nonsynonymous and synonymous substitution of SARS-CoV-2, SARS-CoV, and MERS-CoV genome. *Genomics Inform* 19:e15.
- Spielman SJ, Weaver S, Shank SD, Magalis BR, Li M and Pond SLK (2019) Evolution of viral genomes: Interplay between selection, recombination, and other forces. In: Anisimova M (eds) *Evolutionary Genomics. Methods in Molecular Biology*. Humana Press, New York, NY, vol. 1910.
- Starr TN and Thornton JW (2016) Epistasis in protein evolution. *Protein Sci* 25:1204-1218.
- Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z *et al.* (2020) On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev* 7:1012-1023.
- Taylor L (2021) Covid-19: Is Manaus the final nail in the coffin for natural herd immunity? *BMJ* 372:n394.
- Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N *et al.* (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. DOI: 10.1101/2020.12.21.20248640.
- Usmanova DR, Ferretti L, Povolotskaya IS, Vlasov PK and Kondrashov FA (2015) A model of substitution trajectories in sequence space and long-term protein evolution. *Mol Biol Evol* 32:542-554.
- Van Valen L (1973) A new evolutionary law. *Evol Theor* 1:1-30.
- Wagner GP and Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* 12:204-213.
- Wan Y, Shang J, Graham R, Baric RS and Li F (2020) Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *J Virol* 94:e00127-20.
- Wang M, Li M, Ren R, Li L, Chen E-Q, Li W and Ying B (2020) International expansion of a Novel SARS-CoV-2 Mutant. *J Virol* 94:e00567-20.

- Wang W, Zhao H and Han GZ (2020) Host-virus arms races drive elevated adaptive evolution in viral receptors. *J Virol* 94:e00684-20.
- Weinreich DM, Delaney NF, De Pristo MA and Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111-114.
- Wertheim JO, Chu DKW, Peiris JSM, Pond SLK and Poon LM (2013) A case for the ancient origin of coronaviruses. *J Virol* 87:7039-7045.
- Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Lambson BE, de Oliveira T, Vermeulen M, van der Berg K *et al.* (2021) SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med* 27:622-625.
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J *et al.* (2020) Genome composition and divergence of the Novel Coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27:325-328.
- Xiao J, Fang M, Chen Q and He B (2020) SARS, MERS and COVID-19 among healthcare workers: A narrative review. *J Infect Public Health* 13:843-848.
- Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, Wang Y, Hong Q, Wang S, Zhao Q *et al.* (2021) Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci Adv* 7:eabe5575.
- Yan R, Zhang Y, Li Y, Xia L, Guo Y and Zhou Q (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367:1444-1448.
- Yi K, Kim SY, Bleazard T, Kim T, Youk J and Ju YS (2021) Mutational spectrum of SARS-CoV-2 during the global pandemic. *Exp Mol Med* 53:1229-1237.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME and Fouchier RAM (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814-1820.
- Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, Boerwinkle E and Fu YX (2004) Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol* 4:21.
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang C-L *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270-273.
- Faria NR, Morales Claro I, Candido D, Moyses Franco LA, Andrade PS, Coletti TM, Silva CA, Sales FC, Manuli ER, Aguiar RS *et al.* (2021a) Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: Preliminary findings - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology – Virological, <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manauas-preliminary-findings/586> (accessed 8 Nov 2021).
- Fluxus Technology (2020) Network. v10.2, <https://www.fluxus-engineering.com/index.htm> (accessed 8 Nov 2021).
- Naveca F, da Costa C, Nascimento V, Souza V, Corado A, Nascimento F, Costa Á, Duarte D, Silva G, Mejía M *et al.* (2021) SARS-CoV-2 reinfection by the new Variant of Concern (VOC) P.1 in Amazonas, Brazil - SARS-CoV-2 coronavirus/ nCoV-2019 Genomic Epidemiology – Virological, <https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596> (accessed 8 Nov 2021).
- NERVTAG (2021) NERVTAG paper on COVID-19 variant of concern B.1.1.7 - GOV.UK, <https://www.gov.uk/government/publications/nervtag-paper-on-covid-19-variant-of-concern-b117> (accessed 8 Nov 2021).
- Pond S (2020) Natural selection analysis of global SARS-CoV-2/ COVID-19 enabled by data from GISAID, <https://observablehq.com/@spond/revised-sars-cov-2-analytics-page> (accessed 8 Nov 2021).
- Reuters (2021) Delta COVID variant becoming globally dominant, WHO official says, <https://www.reuters.com/world/delta-covid-variant-becoming-globally-dominant-says-who-official-2021-06-18/> (accessed 8 Nov 2021).
- World Health Organization (2003) Consensus document on the epidemiology of severe acute respiratory syndrome (SARS), Department of Communicable Disease Surveillance and Response, <https://www.who.int/csr/sars/en/WHOconsensus.pdf> (accessed 8 Nov 2021).
- World Health Organization (2020) MERS Situation Update, <https://applications.emro.who.int/docs/WHOEMCSR326E-eng.pdf?ua=1> (accessed 8 Nov 2021).
- World Health Organization (2021a) Weekly epidemiological update on COVID-19 - 3 August 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---3-august-2021>. (accessed on 5 August, 2021).
- World Health Organization (2021b) Weekly epidemiological update on COVID-19 - 13 July 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-july-2021>. (accessed on 14 July, 2021).
- World Health Organization (2021c) Weekly epidemiological update on COVID-19 - 25 May 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-may-2021>. (accessed on 27 May, 2021).
- World Health Organization (2021d) Weekly epidemiological update on COVID-19 - 29 June 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---29-june-2021>. (accessed on 1 July, 2021).
- World Health Organization (2021e) Tracking SARS-CoV-2 variants, <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants> (accessed August 11, 2021).
- World Health Organization (2021f) Weekly epidemiological update on COVID-19 - 1 June 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---1-june-2021>. (accessed on 5 June, 2021).
- World Health Organization (2021g) Weekly epidemiological update on COVID-19 - 27 April 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---27-april-2021>. (accessed on 2 May, 2021).
- World Health Organization (2021h) Weekly epidemiological update on COVID-19 - 7 December 2021, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---7-december-2021>. (accessed on 8 December, 2021).

## Internet Resources Section

- BBC News Mundo (2021) Ómicron: Qué es laepistasis y por qué es la clave para entender qué tan peligrosa es la nueva variante del coronavirus, <https://www.bbc.com/mundo/noticias-59554525> (accessed 7 Dec 2021).
- de Almeida LGP, Lamarca AP, da Silva FJR, Cavalcante L, Gerber AL, Guimarães AP de C, Terra Machado D, Alves C, Mariani D, Felix Cruz T *et al.* (2021) Genomic Surveillance of SARS-CoV-2 in the State of Rio de Janeiro, Brazil: technical briefing - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology - Virological, <https://virological.org/t/genomic-surveillance-of-sars-cov-2-in-the-state-of-rio-de-janeiro-brazil-technical-briefing/683> (accessed 8 Nov 2021).
- ECDC. European Centre for Disease Prevention and Control (2021a) COVID-19 situation update worldwide, as of week 30, updated 5 August 2021, <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases> (accessed 8 Nov 2021).
- ECDC. European Centre for Disease Prevention and Control (2021b) Risk Assessment: Risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA – first update, <https://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-spread-new-variants-concern-eueca-first-update> (accessed 8 Nov 2021).

## Supplementary Material

The following online material is available for this article:

Table S1 – Results for positive selection tests (FUBAR, MEME and FEL).

Table S2 – Results for negative selection tests (FUBAR and FEL).

Table S3 – P.1.1 and P.1.2 sequences in genes/ORFs with sites under positive selection.

Figure S1 – Networks: a) Spike; b) ORF3; c) ORF1a; d) ORF1b.

Figure S2 – Geographical dispersion of lineages P.1.1 (A) and P.1.2(B).

*Associate Editor: Carlos F.M. Menck*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.