Invest Clin 62(Suppl. 2): 18 - 26, 2021 https://doi.org/10.22209/IC.v62s2a02

Importance of mutations in amino acid 484 of the Spike protein of SARS-CoV-2: rapid detection by restriction enzyme analysis

Rossana C Jaspe[1], Yoneira Sulbarn[1], Carmen L Loureiro[1], Pierina D´Angelo[2], Lieska Rodríguez[2], Domingo J Garzaro[1], Héctor R Rangel[1] and Flor H Pujol[1]

[1]Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela.
[2]Instituto Nacional de Higiene "Rafael Rangel", Caracas, Venezuela.

Key words: COVID-19; SARS-CoV-2; Variants of Concern (VOC); RFLP; rapid screening; mutations.

Abstract. Variants of Concern of SARS-CoV-2 (VOCs), the new coronavirus responsible for COVID-19, have emerged in several countries. Mutations in the amino acid 484 of the Spike protein are particularly important and associated with some of these variants: E484K or E484Q. These mutations have been associated with evasion to neutralizing antibodies. Restriction enzyme analysis is proposed as a rapid method to detect these mutations. A search on GISAID was performed in April 2021 to detect the frequency of these two mutations in the sequence available and their association with other lineages. E484K, present in some VOCs, has emerged in several other lineages and is frequently found in recent viral isolates. A small amplicon from the Spike gene was digested with two enzymes: HpyAV, and MseI. The use of these two enzymes allows the detection of mutations at position 484, and to differentiate between these three conditions: non-mutated, and the presence of E484K or E484Q. A 100% correlation was observed with sequencing results. The proposed methodology, which allows for the screening of a great number of samples, will probably help to provide more information on the prevalence and epidemiology of these mutations worldwide, to select the candidates for whole-genome sequencing.

Corresponding author: Flor H Pujol. Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela. Tel/Fax: +58-2125041623. E-mail: fhpujol@gmail.com

Importancia de las mutaciones en el amino acido 484 de la Espiga del SARS-CoV-2: identificación rápida por análisis con enzimas de restricción

Invest Clin 2021; 62 (Suppl. 2): 18-26

Palabras clave: COVID-19; SARS-CoV-2; Variantes de preocupación (VOC); RFLP; detección rápida; mutaciones.

Resumen. En varios países han surgido variantes de preocupación del SARS-CoV-2 (VOC), el nuevo coronavirus responsable de la COVID-19. Las mutaciones en el aminoácido 484 de la proteína de la Espiga (S) son particularmente importantes y están asociadas con algunas de estas variantes: E484K o E484Q. Estas mutaciones han sido asociadas a evasión de la respuesta de anticuerpos neutralizantes. Se propone el análisis con enzimas de restricción como un método rápido para detectar estas mutaciones. En abril de 2021 se realizó una búsqueda en GISAID para detectar la frecuencia de estas dos mutaciones en la secuencia disponible y su asociación con otros linajes. La mutación E484K, presente en algunas VOCs, ha surgido en varios otros linajes y se encuentra con mayor frecuencia en aislados virales recientes. Se generó un producto amplificado de un fragmento pequeño del gen S, que fue digerido con dos enzimas: HpyAV y MseI. El uso de estas dos enzimas permite detectar la mutación en la posición 484 y diferenciar entre estas 3 condiciones: no mutado y la presencia de E484K o E484Q. Se observó una correlación del 100% con los resultados de la secuenciación. La metodología propuesta, que permite el cribado de un gran número de muestras, probablemente ayudará a proporcionar más información sobre la prevalencia y epidemiología de estas mutaciones en todo el mundo, para seleccionar los candidatos para la secuenciación del genoma completo.

Received: 25-06-2021 Accepted: 10-07-2021

INTRODUCTION

In March 2020, the COVID-19 pandemic was declared, caused by an emerging coronavirus, SARS-CoV-2. One year and a half later, this infection has caused almost 200 million cases and almost 3 million deaths worldwide. This virus belongs to the family *Coronaviridae*, order *Nidovirales*. These viruses possess an envelope that surrounds a helicoidal nucleocapsid which packs a large continuous RNA genome of almost 30,000 nt, which codes for 4 structural proteins (nucleocapsid or N, spike or S, matrix or M and envelope or E), 16 non-structural protein and 8 additional accessory ones (Fig. 1). This viral order is unique among the RNA viruses since the viruses belonging to this order encode for an exonuclease, which enables proof-reading capacity to the replication machinery, limiting mutational events. However, the tremendous number of replication events that this virus has undergone, in addition to an elevated frequency of recombination, and to the probable action of host deaminases on the viral genome (1), has allowed the emergence of many point mutations and frequent deletions in the viral genome (2).

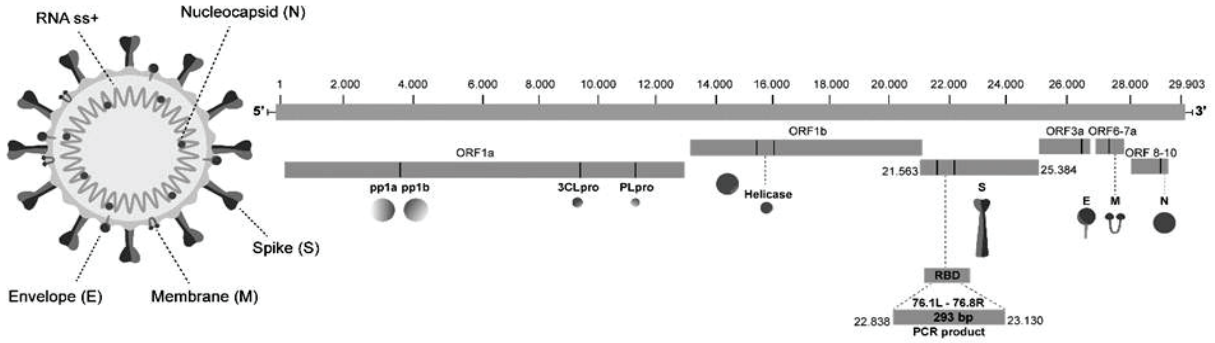

Fig. 1. SARS-CoV-2 virion structure, genome. Structural proteins of SARS-CoV-2 are shown in the virion on the left side of the figure. The genome of almost 30,000 nt is shown on the right side, with the different Open Reading Frames (ORFs) and the proteins produced. RBD is shown inside the S protein, as well as the small PCR-amplified product used for restriction analysis.

Different variants (groups of viruses sharing particular types of mutations) have emerged at the end of 2020. Some of these variants have been defined of Interest (VOI) or Concern (VOC) by WHO: VOI harbor mutations that may confer to these groups of viruses more transmissibility, or partial resistance to protective immunity or treatment, among others. When some of these characteristics are demonstrated for a particular VOI, this VOI is named VOC (3). The variant B1.1.7 (VOC α), for which an increase in incidence after its emergence was observed in the UK, variant B.1.351 (VOC β), with an increase in prevalence in South Africa, variants B.1.1.28.1 or P.1 (VOC γ) and B.1.1.28.2 or P.2 (VOI), which are now predominant in Brazil, and variants B.1.617.1 (VOI θ) and B.1.617.2 (VOC δ) are examples of these variants. The first name corresponds to the lineage according to Pango lineages classification (https://cov-lineages.org/lineages.html) and the Greek letter corresponds to WHO classification (4-7). Genomic surveillance has been proposed for monitoring the introduction of SARS-CoV-2 Variants of Concern (VOCs) in different countries (8,9).

Some of the mutations harbored by these variants are of public health concern for two main reasons: mutation N501Y (tyrosine substituting an asparagine), for example, might lead to increased transmissibility of the viruses harboring this mutation, and mutation E484K (lysine substituting a glutamic acid) has been frequently associated with reinfection cases, and might reduce the neutralizing activity of antibodies produced by vaccination (7,10). Both mutations are located in the Receptor Binding Domain (RBD) of the S protein (Fig. 1). Amino acid 484 of S is of particular interest, since at least two important mutations have emerged in this position: E484K and E484Q. E484K is found in VOCs β and γ, and sometimes in VOC α, while E484Q is found in VOI κ.

A rapid method for detecting those mutations might be useful for screening large quantities of samples, in settings where sequencing facilities are limited. This would be very useful for rapid and massive screening of these variants. In this study, we propose a simple restriction enzyme digestion analysis for the detection of these two mutations in the SARS-CoV-2 Spike: E484K and E484Q.

METHODOLOGY

Analysis of sequences available at GISAID for E484K and E484Q mutations

Sequences available at GISAID in April 2021 were analyzed for the presence of E484K, E484Q and N501Y, at https://

www.gisaid.org/phylodynamics/global/nextstrain/ and https://www.epicov.org/.

RT-PCR

This study was approved by the Bioethical Committee of IVIC. We previously detected in Venezuela, the circulation of variants of the B.1.1.28.1 lineage (P1-like) and samples harboring the E484K mutation without the N501Y one, by sequencing part of the Spike gene in Venezuelan isolates (Jaspe, RC, in preparation). RNA from clinical samples positive by qRT-PCR (classified as wild-type, WT, or harboring E484K) was amplified with primers 76.1L (5´-CCAGATGATTTTACAGGCTGCG-3´) and 76.8R (5´-GTTGCTGGTGCATGTAGAAGTTC-3´) (Fig. 1), using SuperScript III One-Step RT-PCR System with Platinum Taq High Fidelity DNA Polymerase (Thermo Fisher Scientific), and the following PCR conditions: an incubation at 55°C for 30 min, followed by 94°C/3min and 40 cycles of 94°C/15 sec, 55°C/30 sec and 68°C/30 sec, with a final extension of 68°C for 7 min. Superscript II One-Step PCR System with Platinum Taq DNA Polymerase (Thermo Fisher Scientific), was also used successfully.

Restriction analysis

Five μl of the amplicon were digested with 1 unit of HpyAV or MseI for 1 hour at 37°C and then loaded in a 3% agarose gel electrophoresis for band visualization with Ethidium bromide. Restriction results were compared with the sequence obtained by sending PCR purified fragments to Macrogen Sequencing Service (Macrogen, Korea).

RESULTS

A total of 1,224,815 sequences of SARS-CoV-2 were available on April 24 at GISAID. Fig. 2 shows the frequency of sequences harboring the N501Y, E484K, or both mutations. However, this frequency does not reflect the true prevalence of these mutations worldwide, since a strong bias exists, which has been accentuating over the last months. There are strong differences in genomic sequencing capacities between developed and developing countries, and even between developed countries (Table I).

For example, UK is the country that has provided the highest number of sequences to GISAID, more than the USA, even if the total number of COVID-19 cases is 7.4 times lower than the ones in the USA. VOC of the lineage B.1.1.7, which generally lacks the E484K mutations, emerged in the UK and is also frequent in the USA. In contrast, the other two VOCs, which harbor E484K in addition to N501Y, emerged in Brazil and South Africa, countries with robust sequencing capacities but not comparable to the countries previously mentioned (Table I). This is the main reason for the higher abundancy of N501Y mutation, compared to E484K one

TABLE I
NUMBER OF SEQUENCES HARBORING E484K OR N501Y MUTATIONS IN SELECTED COUNTRIES.

Country	COVID-19 cases	Total sequences	N501Y	E484K
UK	4,403,170	382,862	212,371	1,398
USA	32,788,341	331,109	57,661	16,443
Germany	3,277,661	64,503	37,412	1,668
Brazil	14,308,215	5,827	1,063	1,872
South Africa	1,574,370	5,075	1,974	2,020

Total COVID-19 cases and sequences available at GISAID on April 24, 2021. The number of sequences with the N501Y mutation is shown for comparison.

(Fig. 2). E484Q mutation is less commonly found in SARS-CoV-2 isolates (4966 in total until April 24, 2021). Even if the frequency of the E484K mutation appears relatively low in the sequences available at GISAID, this mutation has been found associated with many lineages (Table II).

Fig. 3 shows the restriction pattern of Wild Type (WT) samples and isolates harboring mutations E484K. Digestion with HpyAV of the 293 bp amplicon of the Spike partial genomic region yielded two bands of 170 and 123 bp for samples with E484, while the digestion site is abolished with K484 or Q484. Digestion with MseI yielded a fragment of 183 and 110 for WT samples and Q484, and a fragment of 174, 110, and 9 bp (this last one not observed in the gel) for samples with K484. The presence of E484Q mutation generates a hybrid pattern between the two other situations: no digestion with HpyAV (as E484K) and fragments of 183 and 110 as WT samples when digested with MseI (Fig. 3A). The 9 bp difference of the product between the WT and E484K mutated samples was easily differentiated in 3% agarose gels (Fig. 3B). NuSieve agarose or polyacrylamide gel electrophoresis can also be used for more discrimination, particularly of bands of 174 and 183 bp. However, this was not necessary in our hands.

A total of 40 samples (12 WT and 28 with E484K) were analyzed for their restriction pattern and compared to the presence or not of the mutation in their sequence. A 100% correlation was observed in the detection of the mutations between the two methods (data not shown). No sample was available harboring the E484Q mutation. However, the analysis of the sequence with the E484Q mutation allows inferring a distinct restriction pattern compared to WT and E484K (Fig. 3B).

DISCUSSION

SARS-CoV-2 variants are emerging and spreading rapidly in several parts of the world (3). Mutations E484K and E484Q have been associated with many of the VOCs and may appear in other variants already unknown. Indeed, a survey of sequences available at GIASID showed that these mutations, particularly E484K, have been found in sequences for many lineages, suggesting that this mutation is emerging frequently (Jaspe, R.C. et al., in preparation) (11). As stated

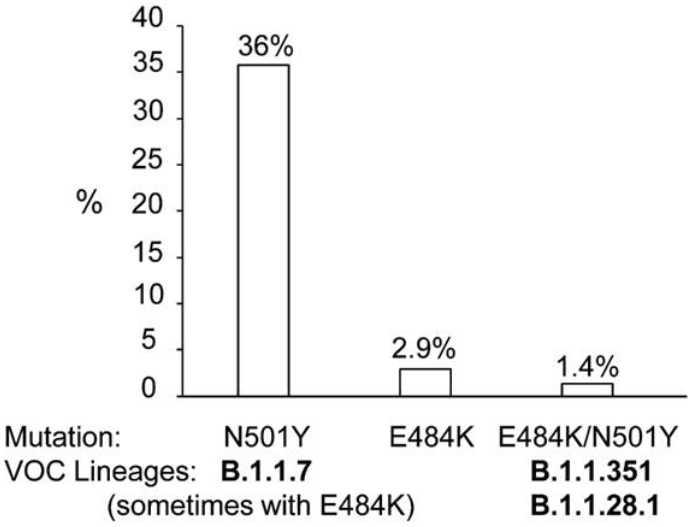

Fig. 2. Frequency of sequences in GISAID (until April 2021) harboring N501Y, E484K mutations, or both. The frequency of N501Y isolates is shown for comparison.

TABLE II
LINEAGES WHERE E484K MUTATION CAN BE FOUND

Mutation	Lineages where some or all sequences harbor this mutation
E484K	Total lineages: 127
Lineages with less than 10 sequences with the mutation (n=99)	A A.23 B B.1.1.1 B.1.1.10 B.1.1.101 B.1.1.115 B.1.1.16 B.1.1.161 B.1.1.214 B.1.1.216 B.1.1.220 B.1.1.241 B.1.1.25 B.1.1.270 B.1.1.273 B.1.1.297 B.1.1.316 B.1.1.317 B.1.1.322 B.1.1.34 B.1.1.348 B.1.1.351 B.1.1.365 B.1.1.38 B.1.1.393 B.1.1.398 B.1.1.406 B.1.1.412 B.1.1.434 B.1.1.461 B.1.1.482 B.1.1.486 B.1.1.51 B.1.1.515 B.1.1.519 B.1.1.57 B.1.110 B.1.111 B.1.131 B.1.134 B.1.139 B.1.160.26 B.1.177.10 B.1.177.12 B.1.177.31 B.1.177.4 B.1.177.40 B.1.177.44 B.1.177.51 B.1.177.62 B.1.177.77 B.1.177.85 B.1.177.86 B.1.214.3 B.1.218 B.1.220 B.1.221 B.1.234 B.1.237 B.1.241 B.1.243 B.1.258 B.1.265 B.1.279 B.1.281 B.1.311 B.1.313 B.1.314 B.1.349 B.1.36 B.1.36.10 B.1.361 B.1.369 B.1.384 B.1.395 B.1.396 B.1.409 B.1.411 B.1.416 B.1.429 B.1.441 B.1.486 B.1.526.1 B.1.526.2 B.1.527 B.1.538 B.1.555 B.1.564 B.1.573 B.1.577 B.1.595 B.1.609 B.1.612 B.1.620 C.11 C.22 L.3 P.1.1
Lineages with 10-100 sequences (n=18)	A.23.1 B.1.1 B.1.1.28 B.1.1.306 B.1.1.318 B.1.1.345 B.1.110.3 B.1.160 B.1.177 B.1.177.88 B.1.243.1 B.1.375 B.1.400 B.1.596 B.1.618 N.9 P.3 R.2
Lineages with more than 100 sequences (n=10)	B.1.2 (104) B.1.619 (167) B.1.1.207 (190) **B.1.1.7*** (378) B.1 (403) P.2 (1.896) R.1 (2.175) **P.1** (4.192) B.1.526 (6.011) **B.1.351** (7.595)

The list of lineages was assessed in the GISAID. The lineages in bold refer to VOCs. The number of lineages harboring the mutation is shown under parenthesis. This analysis includes fewer sequences than the ones included in Fig. 2, since only sequences with high coverage were analyzed. *Some isolates belonging to the B.1.1.7 have acquired the E484K mutation.

before, the E484K mutation has been found in cases of reinfection and may represent a mutation that emerged to escape the presence of neutralizing antibodies (11,12). The frequency of variants or mutations cannot be estimated with the sequences available at GISAID, since huge disparities exist between the sequencing capacities among countries.

The presence of any of these mutations does not necessarily represent that a VOC is circulating in a specific country. Their presence does not necessarily mean that this isolate will gain the phenotypic advantages provided by these mutations in VOCs. Many other mutations present in these VOCs (point mutations and deletions) might be contributing to the enhanced transmissibility of these VOCs (13). However, the specific detection of these two mutations, which play a key role in determining the phenotype of the isolate, might be of particular interest. Thus a rapid method for identifying those mutations should be very useful, particularly in settings where massive whole genome sequencing is not available. Rapid methodologies might be used for the rapid screening of several samples. The whole protocol can be run in a day.

The proposed methodology allows analyzing a great number of samples to select samples that may harbor mutations of concern, before proceeding to whole genome sequencing. On the other hand, once the presence of the variants is confirmed by whole genome sequencing, this method can be used for the rapid estimation of their prevalence in different geographical regions.

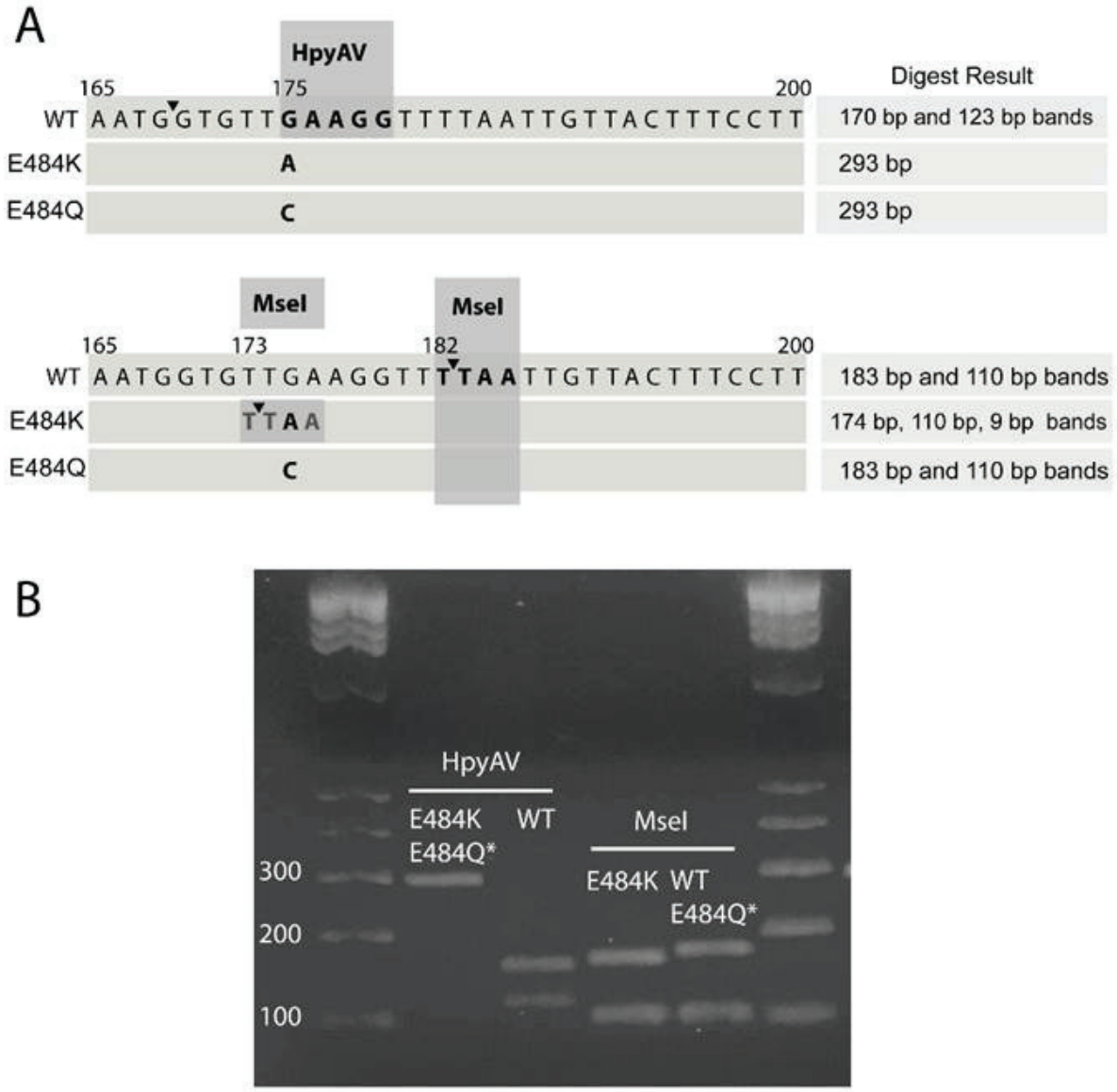

Fig. 3. Restriction analysis of amplicons with E484K or E484Q mutations. A. Sequence of the amplified product showing the restriction sites which discriminate Wild-type (WT) or mutant (E484K or E484Q) viruses. Predicted size of the restriction fragments produced by digestion with each enzyme and visible on agarose gel. The use of these two enzymes generates a restriction pattern characteristic for each situation (WT, E484K and E484Q). The arrow indicates the position in bp of the restriction site. The numbers in the alignment indicate the bp position in the PCR-amplified product (see Fig. 1). Nucleotides 175-177 codes for the amino acid E484 (GAA), K484 (AAA), or Q484 (CAA). B. Agarose gel electrophoresis of digested PCR-amplified products. The PCR-amplified products digested with the enzyme (HpyAV or MseI) were run with molecular weight markers (1Kb plus DNA ladder): smaller bands are signaled (100, 200, and 300 bp). *Samples with E484Q mutations were not available for analysis but the predicted restriction pattern is shown.

ACKNOWLEDGEMENTS

This study was supported by the Ministerio del Poder Popular de Ciencia, Tecnología e Innovación of Venezuela.

REFERENCES

1. **Pujol FH, Zambrano JL, Jaspe R, Loureiro CL, Vizzi E, Liprandi F, Rangel HR**. Biología y evolución del coronavirus causante de la COVID-19. Rev Soc Venezol Microbiol 2020; 40:63-70.
2. **Phan T**. Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol 2020; 81:104260. Available from: *https://doi.org/10.1016/j.meegid.2020.104260.*
3. **Pujol FH, Esparza, J**. COVID-19: virus, variantes y vacunas. Bol Acade Cienc Fis Matemat Natur 2020; LXXXI: 1-10.
4. **Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waites W, Wong KLM, van Zandvoort K, Silverman JD, CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ.** Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science 2021; 9:372(6538):eabg3055. *https://doi.org/10.1126/science.abg3055.*
5. **Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail , Mohale T, Glass AJ, Engelbrecht A, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T.** Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. 2021; 592(7854):438-443. *https://doi.org/10.1038/s41586-021-03402-9.*
6. **Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT, Hulswit RJG, Franco LAM, Ramundo MS, de Jesus JG, Andrade PS, Coletti TM, Ferreira GM, Silva CAM, Manuli ER, Pereira RHM, Peixoto PS, Kraemer MUG, Gaburo N Jr, Camilo CDC, Hoeltgebaum H, Souza WM, Rocha EC, de Souza LM, de Pinho MC, Araujo LJT, Malta FSV, de Lima AB, Silva JDP, Zauli DAG, Ferreira ACS, Schnekenberg RP, Laydon DJ, Walker PGT, Schlüter HM, Dos Santos ALP, Vidal MS, Del Caro VS, Filho RMF, Dos Santos HM, Aguiar RS, Proença-Modena JL, Nelson B, Hay JA, Monod M, Miscouridou X, Coupland H, Sonabend R, Vollmer M, Gandy A, Prete CA Jr, Nascimento VH, Suchard MA, Bowden TA, Pond SLK, Wu CH, Ratmann O, Ferguson NM, Dye C, Loman NJ, Lemey P, Rambaut A, Fraiji NA, Carvalho MDPSS, Pybus OG, Flaxman S, Bhatt S, Sabino EC.** Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. Science 2021; 3;2021.02.26.21252554. *https://doi.org/10.1126/science.abh2644.*
7. **Davies NG, Jarvis CI, CMMID COVID-19 Working Group, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH**. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature 2021; 593:270–274. *https://doi.org/10.1038/s41586-021-03426-1.*
8. **World Health Organization**. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-june-2021.
9. **European Centre for Disease Prevention and Control.** SARS-CoV-2 variants of concern as of 18 June 2021. https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea
10. **Altmann DM, Boyton RJ, Beale R.** Immunity to SARS-CoV-2 variants of concern. Science 2021; 371(6534):1103-1104. *https://doi.org/10.1126/science.abg7404.*
11. **Ferrareze PAG, Franceschi VB, Mayer AM, Caldana GD, Zimerman RA, Thompson CE.** 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. 2021;93:104941. *https://doi.org/10.1016/j.meegid.2021.104941.*

12. **Hoffmann M, Arora P, Groß R, Seidel A, Hörnich BF, Hahn AS, Krüger N, Graichen L, Hofmann-Winkler H, Kempf A, Winkler MS, Schulz S, Jäck HM, Jahrsdörfer B, Schrezenmeier H, Müller M, Kleger A, Münch J, Pöhlmann S.** SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. Cell 2021; 184(9):2384-2393.e12. *https://doi.org/10.1016/j.cell.2021.03.036.*

13. **Khan A, Zia T, Suleman M, Khan T, Ali SS, Abbasi AA, Mohammad A, Wei DQ**. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. J Cell Physiol [On line] 2021 [Cited 2021 Mar 23].: *https://doi.org/10.1002/jcp.30367*.