Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Infection, Genetics and Evolution 97 (2022) 105175

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Codon usage divergence in Delta variants (B.1.617.2) of SARS-CoV-2

Gun Li *, Liang Zhang, Pei Xue

Laboratory for Biodiversity Science, Department of Biomedical Engineering, School of Electronic Information Engineering, Xi'An Technological University, Xi'An, China

ARTICLE INFO

Keywords:
SARS-CoV-2
Delta variants (B.1.617.2)
Biodiversity
Codon usage divergence
Gene evolution

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spreads all over the world and brings great harm to humans in many countries. Many new SARS-CoV-2 variants appeared during its transmission. In the present study, the Delta variants (B.1.617.2) of SARS-CoV-2, which have appeared in many countries, were considered for analysis. In order to evaluate the evolutionary divergence of the Delta variants(B.1.617.2), the codon usage divergence in Delta variants (B.1.617.2) of SARS-CoV-2 was compared to that of the SARS-CoV-2 genomes emerged before June 2020. All Delta variants (B.1.617.2) and 350 early genomes of SARS-CoV-2 in the NCBI database were downloaded. Codon usage pattern including the basic composition, the GC ratio of the third position (GC3) and the first two positions (GC12) in codons, overall GC contents, the effective number of codons (ENC), the codon bias index (CBI), the relative synonymous codon usage (RSCU) values, *etc.*, of all concerned important gene sequences were all calculated. Codon usage divergence of them was calculated *via* summing their standard deviations. The results suggested that base compositions in both Delta variants (B.1.617.2) of SARS-CoV-2 and the early SARS-CoV-2 genomes were similar to each other. However, the internal codon usage divergence for most genes in Delta variants (B.1.617.2) was significantly wider than that of SARS-CoV-2. The RSCU values were further used to explore the synonymous and non-synonymous mutations in the sequences of the Delta variants (B.1.617.2), and the results showed the synonymous mutations are more obvious than the non-synonymous in the concerned sequences. The related codon usage divergence analysis is helpful for further study on the adaptability and disease prognosis of the SARS-CoV-2 variants.

1. Introduction

SARS-CoV-2 was identified as a new pathogen in China in December 2019, and now has been spreading globally for nearly two years. For rapid characterization of SARS-CoV-2, Centers for Disease Control and Prevention (CDC) has classified these variants of concern (VOC), viewing current variants that are being closely monitored and characterized by federal agencies, and by Pango lineage system (Rambaut et al. 2020) (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html#anchor_1632154493691). Delta variants, emerged as a highly contagious virus and was soon listed as a VOC in CDC classification of emerging variants (Afrin et al., 2022), can be characterized by their Pango Lineage (B.1.617.2 and AY lineages). Meanwhile, there were many other variants of SARS-CoV-2, such as the Alpha *variant* (Zoccola et al., 2020), the Beta *variant* (Cerutti et al., 2021; Daming et al., 2021), the P.1 *variant* (de Siqueira et al., 2021), *etc.* have been found during the transmission process. Tracing the genetic diversity of SARS-CoV-2 has attracted the attention of many scientists all over the world (Malick and Fernandes, 2021; Kannan et al., 2021; Laiton-Donato et al., 2021). Bioinformatics tools are very helpful for exploring the phenotypic divergence of SARS-CoV-2 Genomes, especially the diversity of special genes (Dilucca et al., 2020). However, the evolutionary divergence within genes from Delta variants (B.1.617.2) was not studied before. In the present communication, the codon usage divergence of genes for Delta variants (B.1.617.2) was studied by summarizing their standard deviations, and the RSCU values in certain gene sequences were further used to study the changes from the amino acid level by comparing the synonymous and non-synonymous mutations in them.

2. Materials and methods

In order to explore the codon usage divergence in Delta variants (B.1.617.2) of SARS-CoV-2, comparative study on the Delta variants (B.1.617.2) and the early SARS-CoV-2 genomes was conducted. All 153 Delta variants (B.1.617.2) genomes in the NCBI database were searched *via* indexing the 'B.1.617.2' and limiting the length from 20,000 to

* Corresponding author.
E-mail address: ligun@xatu.edu.cn (G. Li).

https://doi.org/10.1016/j.meegid.2021.105175
Received 7 September 2021; Received in revised form 27 November 2021; Accepted 1 December 2021
Available online 3 December 2021

Table 1
Codon usage pattern of special genes in both SARS-CoV-2 genomes and in the Delta variants (B.1.617.2) of SARS-CoV-2.

Species	Gene name	A	T	C	G	GC	GC3s	CBI	ENC	GC12
SARS-CoV-2	*ORF1ab* (*n* = 323)	0.302 ± 0.000	0.324 ± 0.000	0.176 ± 0.000	0.199 ± 0.000	0.374 ± 0.000	0.273 ± 0.000	0.263 ± 0.000	44.156 ± 0.010	0.425 ± 0.000
	S (*n* = 350)	0.294 ± 0.000	0.333 ± 0.000	0.189 ± 0.000	0.184 ± 0.000	0.373 ± 0.000	0.267 ± 0.000	0.298 ± 0.001	44.146 ± 0.025	0.427 ± 0.001
	ORF8 (*n* = 349)	0.276 ± 0.000	0.366 ± 0.001	0.175 ± 0.001	0.183 ± 0.000	0.358 ± 0.001	0.262 ± 0.001	0.419 ± 0.009	43.884 ± 0.097	0.406 ± 0.002
	ORF7b (*n* = 339)	0.235 ± 0.000	0.455 ± 0.000	0.182 ± 0.000	0.129 ± 0.000	0.311 ± 0.000	0.341 ± 0.000	0.467 ± 0.000	31.201 ± 0.000	0.295 ± 0.001
	ORF7a (*n* = 346)	0.295 ± 0.000	0.322 ± 0.000	0.216 ± 0.000	0.167 ± 0.000	0.383 ± 0.000	0.279 ± 0.000	0.426 ± 0.001	52.234 ± 0.095	0.434 ± 0.000
	ORF3a (n = 350)	0.272 ± 0.000	0.334 ± 0.001	0.210 ± 0.000	0.185 ± 0.001	0.394 ± 0.001	0.353 ± 0.002	0.320 ± 0.001	51.651 ± 0.141	0.415 ± 0.001
	M (n = 349)	0.256 ± 0.000	0.318 ± 0.000	0.218 ± 0.000	0.208 ± 0.000	0.426 ± 0.000	0.390 ± 0.000	0.327 ± 0.001	54.181 ± 0.083	0.444 ± 0.000
	N (n = 350)	0.317 ± 0.001	0.210 ± 0.000	0.250 ± 0.000	0.222 ± 0.001	0.472 ± 0.001	0.379 ± 0.001	0.288 ± 0.003	52.706 ± 0.156	0.519 ± 0.000
	ORF6 (n = 349)	0.366 ± 0.001	0.355 ± 0.001	0.140 ± 0.001	0.140 ± 0.001	0.280 ± 0.001	0.290 ± 0.001	0.636 ± 0.004	32.935 ± 0.131	0.274 ± 0.001
Delta variants (B.1.617.2) of SARS-CoV-2	*ORF1ab* (*n* = 145)	0.302 ± 0.000	0.324 ± 0.000	0.176 ± 0.000	0.199 ± 0.000	0.374 ± 0.000	0.273 ± 0.001	0.264 ± 0.002	44.160 ± 0.044	0.425 ± 0.000
	S (*n* = 144)	0.295 ± 0.001	0.331 ± 0.002	0.189 ± 0.001	0.185 ± 0.001	0.374 ± 0.001	0.266 ± 0.002	0.286 ± 0.013	44.344 ± 0.142	0.428 ± 0.002
	ORF8 (*n* = 12)	0.276 ± 0	0.367 ± 0	0.173 ± 0.002	0.183 ± 0	0.357 ± 0.002	0.262 ± 0	0.429 ± 0.017	43.928 ± 0.311	0.404 ± 0.002
	ORF7b (*n* = 132)	0.235 ± 0.001	0.456 ± 0.003	0.182 ± 0.000	0.129 ± 0.001	0.311 ± 0.002	0.342 ± 0.006	0.466 ± 0.005	31.065 ± 0.608	0.294 ± 0.005
	ORF7a (*n* = 115)	0.295 ± 0.000	0.324 ± 0.002	0.214 ± 0.002	0.167 ± 0.000	0.381 ± 0.002	0.279 ± 0.002	0.425 ± 0.005	52.585 ± 0.571	0.432 ± 0.003
	ORF3a (*n* = 151)	0.275 ± 0.006	0.330 ± 0.010	0.209 ± 0.001	0.187 ± 0.004	0.396 ± 0.004	0.356 ± 0.003	0.325 ± 0.029	52.657 ± 0.958	0.415 ± 0.005
	M (*n* = 148)	0.256 ± 0.000	0.317 ± 0.000	0.220 ± 0.000	0.208 ± 0.001	0.428 ± 0.001	0.390 ± 0.000	0.331 ± 0.005	54.039 ± 0.240	0.446 ± 0.001
	N (n = 145)	0.318 ± 0.003	0.212 ± 0.002	0.250 ± 0.001	0.220 ± 0.002	0.470 ± 0.001	0.379 ± 0.001	0.275 ± 0.008	53.087 ± 0.907	0.516 ± 0.002
	ORF6 (*n* = 152)	0.366 ± 0.000	0.355 ± 0.000	0.140 ± 0.000	0.140 ± 0.000	0.280 ± 0.000	0.290 ± 0.000	0.636 ± 0.000	32.922 ± 0.000	0.274 ± 0.000

33,000 bases, and downloaded. The early SARS-CoV-2 genomes were searched *via* selecting the ones before June, 1, 2020 in the NCBI database and downloaded. Basic compositions, GC contents including percentage of overall GC, GC12 and GC3, CBI, and ENC of particular genes (*ORF1ab*, *S*, *ORF8*, *ORF7b*, O*RF7a*, *ORF3a*, *M*, *N* and *ORF6*) in both SARS-CoV-2 genomes and the Delta variants (B.1.617.2) were counted or calculated (Gun et al., 2021). The summations of their standard deviations were used to analyze their codon usage divergence. Mutations in coding sequences were further explored by comparing the codon usage level within both Delta variants (B.1.617.2) and normal SARS-CoV-2 genomes, and the amino acid level changes were also studied *via* comparing their differences in non-synonymous rate. The Matlab software was used for extracting the sequences, counting the basic composition and calculating the codon usage pattern for gene

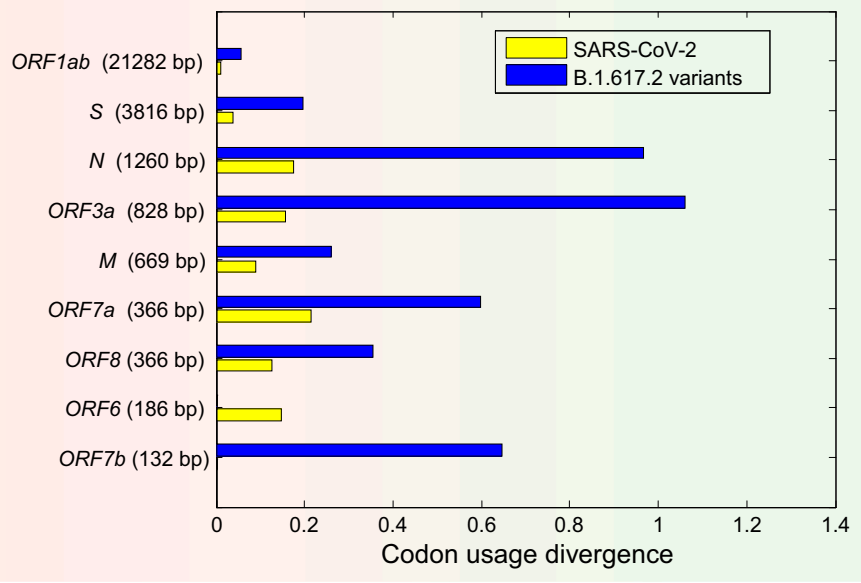

Fig. 1. Codon usage divergence in different sequences in both SARS-CoV-2 genomes and in the Delta variants (B.1.617.2) of SARS-CoV-2.

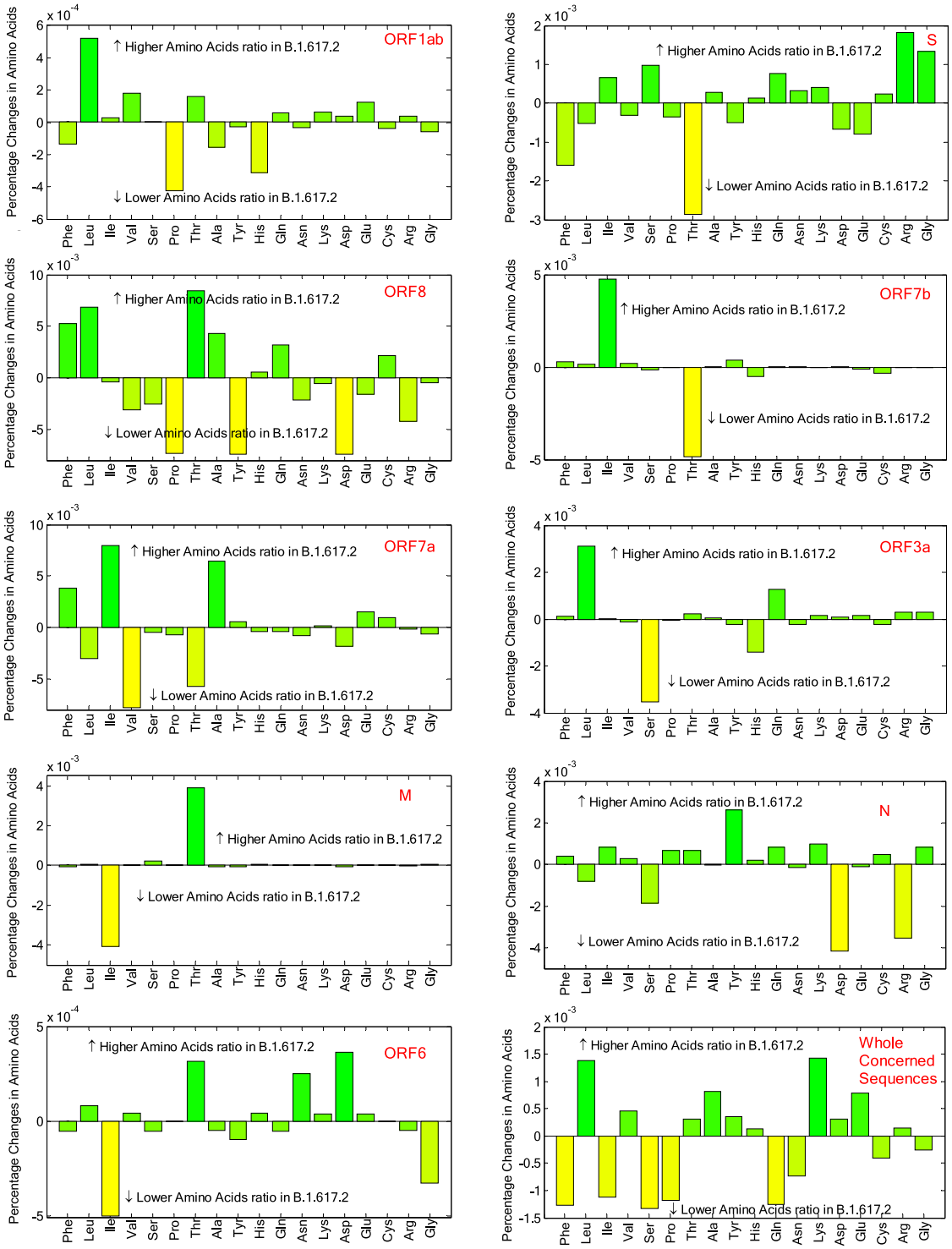

Fig. 2. Changes of proportion in protein *via* amino acids ratio, revealing the non-synonymous codon changes in Delta variants (B.1.617.2).

sequences, and the Excel software was used to calculate mean values, standard deviation of their codon usage patterns.

3. Results

All Delta variants (B.1.617.2) in the NCBI database and the early (before June 1, 2020) SARS-CoV-2 genomes were considered to study their codon usage pattern. The comparative characterization of the codon usage patterns of both SARS-CoV-2 and Delta variants (B.1.617.2) are shown in the Table 1. The results show that all the concerned coding sequences preferred to select A and U as their components, and the base composition of the mutant sequences do not differ significantly from that of the normal sequences. However, the differences of the CBI and ENC between the mutant sequences and the normal sequences are very obvious. From the results, obviously, not all concerned genes exist in every genome, for instance, gene sequence quantity of *ORF8* (only 12 sequences) is much lower than the number of genomes for Delta variants (B.1.617.2). Generally, GC3 ratio and that of GC12 should equal to 0.5

or so, if the evolutionary pressure (including mutation and selection) is equally effect on them. However, in the present study, although the coding sequences prefer to use the A and U as their bases, GC12 ratio is more than GC3 in all sequences except *ORF7b*, both in Delta variants (B.1.617.2) and normal SARS-COV-2.

In order to explore the diversity degree for codon usage patterns in both of the mutant sequences and the original sequences, the summation of deviations was considered to calculate the codon usage divergence of them. The results are shown in the Fig. 1. The abscissa of the graph denotes the codon usage divergence, and the vertical coordinates denotes the gene names and their lengths. From the Fig. 1, all codon usage divergence values of coding gens in mutant sequences are higher than that of normal sequences except the *ORF6*. In particular, the codon usage divergence in *N*, *ORF3a*, *ORF7a* and *ORF7b* in mutant sequences is obviously higher, revealing that these coding sequences are susceptible to mutation. Interestingly, there is no diversity in *ORF6* of Delta variants (B.1.617.2), and the codon usage divergence in *ORF7b* of original SARS-CoV-2 sequences is much lower (=0.002), revealing that these genes are conserved in the genome in which they are located.

The RSCU values, another codon usage pattern parameter of the concerned sequences were further calculated. To explicit the degree of synonymous mutations in the Delta variants (B.1.617.2), the difference of the RSCU values for genes between Delta variants (B.1.617.2) and normal sequences are shown in the Supplementary File.1. In Delta variants (B.1.617.2) sequences, ORF8, ORF7a and ORF3a are the ones with obvious synonymous mutation in codon usage level. Non-snonymous changes in amino acid level for each concerned gene sequences were further counted, and the results are shown in the Fig. 2. The display range values of ordinate for the ORF1ab and the ORF6 are all 10^{-4}, indicating the smaller non-snonymous mutation rate within their coding sequences, which is consist with the main conclusion shown by the Fig. 1.

4. Discussion

According to the CDC, the Delta variants had some potential epidemiological impacts, such as increasing transmissibility, reduction in neutralization by some EUA monoclonal antibody treatments, and reduction in neutralization by post-vaccination sera. The codon usage divergence in virus genomes is crucial for exploring the evolutionary trends, and helpful for molecular treatment (Kirola, 2021; Cao et al., 2021; Onodera et al., 2021). Although diversity of the SARS-CoV-2 has been studied previously, relevant studies on the Delta variants (B.1.617.2) were limited with genomic quantity (Dilucca et al., 2020). Furthermore, previous other researches did not highlight the codon usage divergence of certain genes. In the present study, all 153 Delta variants (B.1.617.2) genomes of SARS-CoV-2 in the NCBI database as of August 20, 2021 were considered to study the codon usage divergence of important genes in them. Most of the B.1.617.2 variant sequences were collected in April and May 2021. Therefore, the sequences of SARS-CoV-2 used in the present were the earlier ones in the NCBI database. For comparison, all SARS-CoV-2 genomes in the NCBI database for those release date before June 1, 2020 were all downloaded for present study. Such sample selection method could ensure the reliability of comparison as much as possible for their relatively similar prevalence temporal spans.

Exploring the viral evolution requires a characterization of the sequences and their changes. Alignment methods, whether base on the nucleic acid sequences or the protein sequences, were used frequently. When function of a virus is concerned, its protein sequences are more significant (Li et al., 2021). However, not all mutated sequences can be inherited stably, it is more significant to study the codon usage divergence in virus sequences to explicit the stability of the viral genes. From the results shown in the Fig. 1, the codon usage divergence of genes in B.1.617.2 variant sequences are higher than those of SARS-CoV-2 except the *ORF6*, revealing that most coding genes in B.1.617.2 variant sequences are more unstable. The *ORF1ab* and *S* sequences are usually used as target genes for testing novel coronavirus because of their stability and proper lengths (Yan et al., 2020); and the codon usage divergence of them (shown in the Fig. 1) are much lower than other coding sequences in Delta variants (B.1.617.2). Further, in the present study, from the amino acid level, the encoding products of *ORF1ab* and *S* sequences also show smaller diversity as shown in the Fig. 2.

The B.1.617.2 is of Delta SARS-CoV-2 variants (Kannan et al., 2021); it has spread all over the world (Rahimi et al., 2021) and caused thousands of deaths daily. The Delta variant has a stable but slightly reorganized receptor-binding interface that can lead its immune evasion (Baral et al., 2021). Understanding on the Delta variants (B.1.617.2) genome, especially its high divergence, is essential for both the development of new treatments and improving the approaches for virus detection (Holmes et al., 2020). However, further researches, such as research on the genetic diversity in Delta variants (B.1.617.2) from certain geographical areas is an essential issue for exploring their distribution. The codon usage divergence of coding sequences in Delta variants (B.1.617.2) of SARS-CoV-2 showed similar characteristic to the constituent diversity at the amino acid level.

In conclusion, codon usage divergence of typical genes in Delta variants (B.1.617.2) is higher than those of early SARS-CoV-2 genomes, revealing that the divergence in the Delta variants (B.1.617.2) would be more substantial. The changes caused by the non-synonymous mutations of the coding sequences in Delta variants (B.1.617.2) should be addressed more systematically in future.

Author contribution

Gun Li conceived and designed the study, analyzed and interpreted the data, and drafted the manuscript. Liang Zhang analyzed the data and revised final version of manuscript. Pei Xue analyzed the data and revised final version of manuscript.

Declaration of competing interest

The author declares that there is no conflict of interests.

Acknowledgment

This research work was supported by Special Scientific Research Plan Project of Education Department of Shaanxi Province (No.18JK0377).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.105175.

References

Afrin, Sultana Zahura, Islam, Md Taohidul, Paul, Shyamal Kumar, Kobayashi, Nobumichi, Parvin, Rokshana, 2022. Dynamics of SARS-CoV-2 variants of concern (VOC) in Bangladesh during the first half of 2021. Virology 565, 29–37.

Baral, Prabin, Bhattarai, Nisha, Hossen, Md Lokman, Stebliankin, Vitalii, Gerstman, Bernard S., Narasimhan, Giri, Chapagain, Prem P., 2021. Mutation-induced changes in the receptor-binding interface of the SARS-CoV-2 Delta variant B.1.617.2 and implications for immune evasion. Biochem. Biophys. Res. Commun. 574, 14–19.

Cao, Canhui, He, Liang, Yuan Tian, Yu, Qin, Haiyin Sun, Ding, Wencheng, Gui, Lingli, Peng, Wu., 2021. Molecular epidemiology analysis of early variants of SARS-CoV-2 reveals the potential impact of mutations P504L and Y541C (NSP13) in the clinical COVID-19 outcomes. Infect. Genet. Evol. 92, 104831.

Cerutti, Gabriele, Rapp, Micah, Guo, Yicheng, Bahna, Fabiana, Bimela, Jude, Reddem, Eswar R., Yu, Jian, Wang, Pengfei, Liu, Lihong, Huang, Yaoxing, Ho, David D., Kwong, Peter D., Sheng, Zizhang, Shapiro, Lawrence, 2021. Structural basis for accommodation of emerging B.1.351 and B.1.1.7 variants by two potent SARS-CoV-2 neutralizing antibodies. Structure 29 (7), 655–663.

Daming, Zhou, Wanwisa, Dejnirattisai, Piyada, Supasa, Liu, Chang, Mentzer, Alexander J., Ginn, Helen M., Zhao, Yuguang, Duyvesteyn, Helen M.E., Tuekprakhon, Aekkachai, Nutalai, Rungtiwa, Wang, Beibei, Paesen, Guido C., Lopez-

Camacho, Cesar, 2021. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell 189, 2348–2361.
de Siqueira, Isadora Cristina, Camelier, Aquiles Assunção, Maciel, Elves A.P., Nonaka, Carolina Kymie Vasques, Neves, Margarida Celia L.C., Macêdo, Yasmin Santos Freitas, de Sousa, Karoline Almeida Félix, Araujo, Victor Costa, Paste, Aurea Angelica, de Freitas Souza, Bruno Solano, Grăf, Tiago, 2021. Early detection of P.1 variant of SARS-CoV-2 in a cluster of cases in Salvador, Brazil. Int. J. Infect. Dis. 108, 252–255.
Dilucca, Maddalena, Forcelloni, Sergio, Georgakilas, Alexandros G., Giansanti, Andrea, Pavlopoulou, Athanasia, 2020. Codon usage and phenotypic divergences of SARS-CoV-2 genes. Viruses 12, 498.
Gun, Li, Liang, Zhang, Pei, Xue, 2021. Codon usage pattern and genetic diversity in chloroplast genomes of Panicum species. Gene 802, 145866.
Holmes, Edward C., O'Toole, Áine, Hill, Verity, McCrone, John T., Ruis, Christopher, du Plessis, Louis, Pybus, Oliver G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407.
Kannan, Saathvik R., Spratt, Austin N., Cohen, Alisha R., Hasan Naqvi, S., Chand, Hitendra S., Quinn, Thomas P., Lorson, Christian L., Byrareddy, Siddappa N., Singh, Kamal, 2021. Evolutionary analysis of the Delta and Delta plus variants of the SARS-CoV-2 viruses. J. Autoimmun. 124, 102715.
Kirola, L., 2021. Genetic emergence of B.1.617.2 in COVID-19. New Microbe New Infect. 43, 100929.
Laiton-Donato, Katherine, Franco-Munoz, Carlos, Alvarez-Díaz, Diego A., Ruiz-Moreno, Hector Alejandro, Usme-Ciro, Jose A., Andr'es Prada, Diego, Reales-Gonzalez, Jhonnatan, et al., 2021. Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. Infect. Genet. Evol. 95, 105038.
Li, Xue, Zhang, Liying, Chen, Si, Ji, Weilong, Li, Chang, Ren, Linzhu, 2021. Recent progress on the mutations of SARS-CoV-2 spike protein and suggestions for prevention and controlling of the pandemic. Infect. Genet. Evol. 93, 104971.
Malick, M.S.S., Fernandes, H., 2021. The genomic landscape of SARS-CoV-2: surveillance of variants of concern. Adv. Mol. Pathol. https://doi.org/10.1016/j.yamp.2021.06.006.
Onodera, Taishi, Shunsuke Kita, Yu, Adachi, Saya Moriyama, Sato, Akihiko, Nomura, Takao, Sakakibara, Shuhei, et al., 2021. A SARS-CoV-2 antibody broadly neutralizes SARS-related coronaviruses and variants by coordinated recognition of a virus vulnerable site. Immunity. https://doi.org/10.1016/j.immuni.2021.08.025.
Rahimi, Azadeh, Mirzazadeh, Azin, Tavakolpour, Soheil, 2021. Genetics and genomics of SARS-CoV-2: a review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. Genomics 113, 1221–1232.
Rambaut, A., Holmes, E.C., O'Toole, Á, et al., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 5, 1403–1407. https://doi.org/10.1038/s41564-020-0770-5.
Yan, C., Cui, J., Huang, L., Du, B., Chen, L., Xue, G., et al., 2020. Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. Clin. Microbiol. Infect. 26, 773–779. https://doi.org/10.1016/j.cmi.2020.04.001.
Zoccola, Roberto, Beltramo, Chiara, Magris, Gabriele, Peletto, Simone, Acutis, Pierluigi, Bozzetta, Elena, Radovic, Slobodanka, Zappulla, Francesco, Porzio, Anna Maria, Gennero, Maria Silvia, Dondo, Alessandro, Pasqualini, Chiara, Griglio, Bartolomeo, Ferrari, Angelo, Giuseppe, Ru, Goria, Maria, 2020. First detection of an Italian human-to-cat outbreak of SARS-CoV-2 Alpha variant - lineage B.1.1.7. One Health 13, 100295.