



HAL
open science

Spreading of a new SARS-CoV-2 N501Y spike variant in a new lineage

Philippe Colson, Anthony Levasseur, Jeremy Delerce, Lucile Pinault, Pierre Dudouet, Christian Devaux, Pierre-Edouard Fournier, Bernard La Scola, Jean-Christophe Lagier, Didier Raoult

► To cite this version:

Philippe Colson, Anthony Levasseur, Jeremy Delerce, Lucile Pinault, Pierre Dudouet, et al.. Spreading of a new SARS-CoV-2 N501Y spike variant in a new lineage. *Clinical Microbiology and Infection*, 2021, 10.1016/j.cmi.2021.05.006 . hal-03299956

HAL Id: hal-03299956

<https://amu.hal.science/hal-03299956>

Submitted on 16 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

TITLE PAGE

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Full-length title:

Spreading of a new SARS-CoV-2 N501Y spike variant in a new lineage

Short title (for the running head): SARS-CoV-2 N501Y spike variant in a new lineage

Author list: Philippe COLSON^{1,2}, Anthony LEVASSEUR^{1,2}, Jeremy DELERCE¹, Lucile PINAULT¹, Pierre DUDOUE¹, Christian DEVAUX^{1,2}, Pierre-Edouard FOURNIER^{1,2}, Bernard LA SCOLA^{1,2}, Jean-Christophe LAGIER^{1,2}, Didier RAOULT^{1,2} *

Affiliations: ¹ IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; ² Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD), Assistance Publique - Hôpitaux de Marseille (AP-HM), MEPHI, 27 boulevard Jean Moulin, 13005 Marseille, France

*** Corresponding author:** Didier Raoult, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: didier.raoult@gmail.com

Key words: SARS-CoV-2; variant; spike; N501Y; emergence; Covid-19; epidemic

Word counts: abstract, 198; text, 1,200

Figures: 2; **Table:** 0; **References:** 15

ABSTRACT

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Objectives: Our surveillance of the SARS-CoV-2 genomic epidemiology led us to detect several variants since summer 2020. We report the recent spread of a new SARS-CoV-2 spike 501Y variant.

Methods: SARS-CoV-2 sequences obtained from human nasopharyngeal samples by Illumina next-generation sequencing were analyzed using Nextclade and an in-house Python script and compared using BLASTn to the GISAID database. Phylogeny was performed using the IQ-TREE software.

Results: We identified that SARS-CoV-2 genomes from four patients diagnosed in our institute harbored a new set of amino acid substitutions including L18F;L452R;N501Y;A653V;H655Y;D796Y;G1219V±Q677H. These spike N501Y genomes are the first of Nextstrain clade 19B. We obtained partial spike gene sequences of this genotype for an additional 43 patients. All patients infected with this genotype were diagnosed since mid-January 2021. We detected 42 other genomes of this genotype in GISAID, which were obtained from samples collected in December 2020 in four cases and in 2021 in 38 cases. The 89 sequences obtained in our institute or other laboratories originated from the Comoros archipelago, Western European countries (mostly metropolitan France), Turkey and Nigeria.

Conclusion: These findings warrant further studies to investigate the spread, epidemiological and clinical features, and sensitivity to immune responses of this variant.

TEXT

48

49

50 **Introduction**

51 SARS-CoV-2 has now spread worldwide for >1 year
52 (<https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>). The emergence
53 of major viral variants expanded dramatically since summer 2020 although this only recently
54 came into the spotlight. Thus, we have implemented since SARS-CoV-2 occurrence in France
55 in late February 2020 a surveillance based on their genomes obtained by next-generation
56 sequencing [1]. During summer 2020 we detected the emergence of 10 viral variants,
57 concomitantly with the re-increase of SARS-CoV-2 diagnoses [2]. These variants have been
58 since then responsible for juxtaposed or successive epidemics and accounted in various
59 proportions over time for the total burden of cases. Since January 1st, 2021, we also performed
60 a systematic screening of variants that predominated or arose in our geographical area with
61 variant-specific in-house qPCR assays ([https://www.mediterranee-infection.com/procedure-](https://www.mediterranee-infection.com/procedure-pour-la-detection-par-qpcr-des-variants-sars-cov-2-marseille-4/)
62 [pour-la-detection-par-qpcr-des-variants-sars-cov-2-marseille-4/](https://www.mediterranee-infection.com/procedure-pour-la-detection-par-qpcr-des-variants-sars-cov-2-marseille-4/); [https://www.mediterranee-](https://www.mediterranee-infection.com/procedure-pour-la-detection-par-qpcr-des-variants-sars-cov-2-n501y/)
63 [infection.com/procedure-pour-la-detection-par-qpcr-des-variants-sars-cov-2-n501y/](https://www.mediterranee-infection.com/procedure-pour-la-detection-par-qpcr-des-variants-sars-cov-2-n501y/)). These
64 variants include the Marseille-4 variant (Nexstrain clade 20A.EU1) [3], which has been the
65 most prevalent one between August 2020 and January 2021, and the rapidly-spreading
66 variants 20I/501Y.V1, 20H/501Y.V2, and 20J/501Y.V3, which harbor in their spike the
67 amino acid (aa) substitution N501Y [4]. We describe herein the first SARS-CoV-2 spike
68 501Y variant belonging to Nextstrain clade 19B.

69

70 **Materials and methods**

71 SARS-CoV-2 genomes were obtained directly from nasopharyngeal swab fluid by
72 next-generation sequencing (NGS) with the Illumina Nextera XT paired-end strategy on a

73 MiSeq instrument (Illumina Inc., San Diego, CA, USA), as previously described [5]. Genome
74 consensus sequences were generated through mapping on the SARS-CoV-2 genome GenBank
75 accession no. NC_045512.2 (Wuhan-Hu-1 isolate) with the CLC Genomics workbench v.7.
76 In addition, sequences of the spike gene 5'-region (nucleotides 1-1,854 in reference to
77 NC_045512.2) were obtained by NGS after PCR amplification with an in-house protocol
78 (Supplementary Material Data). Sequences were analyzed using Nextclade
79 (<https://clades.nextstrain.org/>) [6] and an in-house Python script. Other SARS-CoV-2
80 sequences of this spike 501Y clade 19B genotype were searched using BLASTn [7] among all
81 674,762 sequences available in the GISAID database (<https://www.gisaid.org/>) [8] as on
82 March 3rd, with genomes obtained in our institute as queries. Phylogeny reconstruction was
83 performed using the IQ-TREE software with the GTR Model and 1,000 ultrafast bootstrap
84 repetitions after alignment of genomes using MAFFT v.7; the tree was visualized with the
85 iTOL (Interactive Tree Of Life) software [11,12,21]. SARS-CoV-2 culture was performed by
86 inoculating nasopharyngeal samples on Vero E6 cells, as previously described [9].

87 **Ethics**

88 Data have been generated as part of the routine work at Assistance Publique-Hôpitaux de
89 Marseille (Marseille university hospitals). This study has been approved by the ethics
90 committee of our institution (N°2020-016-03). According to European General Data
91 Protection Regulation No 2016/679, patients were informed of the potential use of their
92 medical data and that they could refuse the use of their data. The analysis of collected data
93 followed the MR-004 reference methodology registered under No. 2020-151 and No. 2020-
94 152 in the AP-HM register.

95

96 **Results**

97 Our SARS-CoV-2 genomic surveillance allowed us to detect in four patients SARS-CoV-2-
98 diagnosed in our institute a variant harboring a new combination of eight mutations in the
99 spike protein. These mutations include aa substitution N501Y, associated with aa
100 substitutions L18F, L452R, A653V, H655Y, D796Y, and G1219V, and a synonymous
101 mutation at nucleotide position 22,468 [in reference to NC_045512.2]; Figures 1a, 1b]; an
102 eighth spike aa substitution, Q677H, was identified in one of these genomes. Other nucleotide
103 or aa substitutions or deletions are present in 14 other proteins; these changes notably include
104 deletions in ORF3a and ORF8 genes (Figure 1a). Overall, 24-31 nucleotide substitutions, 14-
105 19 aa substitutions and three deletions are present in these four SARS-CoV-2 genomes
106 (Figure 1a; GISAID accession numbers: EPI_ISL_1097023; EPI_ISL_1097024;
107 EPI_ISL_1201045; EPI_ISL_1201046). Viral isolates were obtained by culturing the four
108 patients' nasopharyngeal sample, which will allow further analyses of the phenotypic features
109 of this variant. In addition, sequences that span codons 1-618 of the spike gene and allow the
110 identification of this variant by covering 4 of its hallmark mutations (C21614T (substitution
111 L18F), G22468T, T22917G (L452R), A23063T (N501Y)) and by being devoid of the
112 Nextstrain clade 20 A23403G mutation, were obtained for 43 other patients (deposited at:
113 <https://doi.org/10.35081/43r6-sz33>).

114 We first detected in our institute this variant we named Marseille-501 in a patient
115 sampled on January 18th, 2021, and its number increased from week 5 (Supplementary
116 Material Figure S1). During the January 18th-February 23rd period (37 days), SARS-CoV-2
117 genotype was obtained by sequencing for 1,016 patients. We identified in the GISAID
118 database 42 other SARS-CoV-2 genomes, all from humans, with the same spike aa
119 substitution set (Figure 2). Eight originated from Mayotte, a French overseas department in
120 Comoros archipelago. The others originated from France (n=22), Denmark (n=2),
121 Netherlands (n=2), Belgium (n=2), England (n=1), Turkey (n=4), and Nigeria (n=1). The

122 most ancient samples were collected mid-December in Denmark and late December in the
123 Netherlands. All 38 other SARS-CoV-2 genomes were obtained from samples collected in
124 2021 (Supplementary Material Figure S1).

125 Phylogenetic analysis conducted with the four genomes from our institute and the 42
126 other GISAID genomes showed that all 46 sequences were clustered and delineated a strongly
127 supported (bootstrap value: 100%) 19B subclade that stands apart from other clade 19B
128 genomes (Figure 2). Moreover, 21 genomes, one from our institute, the 8 ones from Mayotte,
129 and 12 other genomes from France or Turkey were clustered and all harbored the spike
130 Q677H substitution.

131 The 47 patients we diagnosed as infected with this variant had a mean age of 39 ± 19
132 years (range=16-92); 26 (55%) were male. Epidemiological and/or clinical characteristics
133 were available for 12 patients (Supplementary Table S1): three patients had travelled to or
134 originated from the Comoros, another one originated from Guinea-Conakry but had not
135 travelled abroad recently (Supplementary Table S1). Clinical symptoms were available for 9
136 patients: 4 were asymptomatic and 34 (85%) had mild symptoms including fever, rhinorrhea,
137 cough, headache, asthenia, or myalgia. A 94-year-old male patient developed SARS-CoV-2-
138 associated hypoxemic pneumonia 4 days after the second administration of the
139 Pfizer/BioNTech vaccine and died 7 days later. Another patient with severe obesity was
140 admitted in intensive care unit.

141

142 **Discussion**

143 We report herein the spread of the first SARS-CoV-2 501Y variant belonging to clade
144 19B. Its spike protein harbors several aa substitutions recently reported to emerge in various
145 lineages and/or to be associated with immune escape (L18F, L452R, N501Y, H655Y,
146 Q677H) (Supplementary Material Data). For instance, N501Y is part of several variants that

147 belong to distinct lineages and have been detected worldwide in association with various
148 combinations of mutations [4], while L18F has been reported in 20E strains in England, in a
149 20I/N501Y.V1 substrain with an increased replicative fitness, in 20H/501Y.V2 strains with a
150 faster propagation in presence of convalescent plasma, and in a majority of 20J/501Y.V3
151 variants [10].

152 It is currently unresolved how such strains harboring new blocks of mutations in the
153 spike emerge. Overlooked genome evolution following transmission to minks and back to
154 humans [3], or promotion and selection of mutations during administration of remdesivir or
155 convalescent plasma [11] have been suspected. Also, recombinations known to be common
156 amongst coronaviruses may occur between SARS-CoV-2 strains [12]. Anyway, this new
157 501Y variant is another example that same aa substitutions can occur in distinct lineages and
158 follow convergent evolution [4]. Epidemiologically, this new variant infected patients in
159 diverse countries, but with a predominance in the Comoros archipelago (11/47 documented
160 cases) and in metropolitan France. Based on limited data available, it is unknown if this new
161 variant is more or less transmissible compared to other strains and associated with any
162 particular clinical feature.

163 Overall, these data highlight the need for genomic epidemiology surveillance of
164 SARS-CoV-2 strains. The incidence of this new variant and its epidemiological and clinical
165 features deserve being closely monitored. Moreover, considering its particular association of
166 numerous spike aa substitutions, its sensitivity to neutralizing antibodies and to plasma from
167 convalescent or vaccinated persons should be investigated, which will be done for the isolates
168 cultured in our institute.

169

170

171 **Acknowledgments**

172 We are thankful to Marielle Bedotto and Ludivine Brechard for their technical help. The
173 manuscript has been edited by a native English speaker.

174

175 **Author contributions**

176 Conceived and designed the experiments: PC, DR. Contributed materials/analysis tools: PC,
177 AL, JD, LP, PD, CD, PEF, BLS, JCL. Analyzed the data: PC, AL, JD, JCL, BLS, DR. Wrote
178 the paper: PC, DR.

179

180 **Transparency declaration**

181 This work was supported by the French Government under the “Investments for the Future”
182 program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-
183 IAHU-03 and was also supported by Région Provence Alpes Côte d’Azur and European
184 funding FEDER PRIMMI (Fonds Européen de Développement Régional-Plateformes de
185 Recherche et d’Innovation Mutualisées Méditerranée Infection), FEDER PA 0000320
186 PRIMMI.

187 The authors have no conflicts of interest to declare. Funding sources had no role in the design
188 and conduct of the study; collection, management, analysis, and interpretation of the data; and
189 preparation, review, or approval of the manuscript.

190

191

192

193

REFERENCES

194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217

[1] Colson P, Levasseur A, Delerce J, Chaudet H, Bossi V, Ben Khedher M, et al. Dramatic increase in the SARS-CoV-2 mutation rate and low mortality rate during the second epidemic in summer in Marseille. IHU Preprints 2020. doi: <https://doi.org/10.35088/68c3-ew82>.

[2] Fournier PE, Colson P, Levasseur A, Gautret P, Luciani L, Bedotto M, et al. Genome sequence analysis enabled deciphering the atypical evolution of COVID-19 in Marseille, France. medRxiv 2021. doi: <https://doi.org/10.35088/kmct-tj43>.

[3] Fournier PE, Colson P, Levasseur A, Devaux C, Gautret P, Bedotto M, et al. Emergence and outcome of the SARS-CoV-2 "Marseille-4" variant. IHU pre-prints 2021. doi: <https://doi.org/10.35081/xcrm-6t77>.

[4] Garcia-Beltran WF, Lam EC, Denis KS, Nitido AD, Garcia ZH, Hauser BM, et al. Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. medRxiv 2021. doi: <https://doi.org/10.1101/2021.02.14.21251704>.

[5] Colson P, Levasseur A, Gautret P, Fenollar F, Hoang VT, Delerce J et al. Introduction into the Marseille geographical area of a mild SARS-CoV-2 variant originating from sub-Saharan Africa. Travel Med Infect Dis 2021, 40, 101980.

[6] Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018;34:4121-4123.

[7] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-410.

[8] Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill 2017;22.

- 218 [9] La Scola B, Le Bideau M, Andreani J, Hoang VT, Grimaldier C, Colson P, et al. Viral
219 RNA load as determined by cell culture as a management tool for discharge of SARS-
220 CoV-2 patients from infectious disease wards. *Eur J Clin Microbiol Infect Dis*
221 2020;39:1059-1061.
- 222 [10] Grabowski F, Kochanczyk M, Lipniacki T. L18F substrain of SARS-CoV-2 VOC-
223 202012/01 is rapidly spreading in England. medRxiv 2021 doi:
224 <https://doi.org/10.1101/2021.02.07.21251262>.
- 225 [11] Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, et al. SARS-CoV-2
226 evolution during treatment of chronic infection. *Nature* 2021; Feb 5. doi:
227 [10.1038/s41586-021-03291-y](https://doi.org/10.1038/s41586-021-03291-y). Online ahead of print.
- 228 [12] Varabyou A, Pockrandt C, Salzberg SL, Pertea M. Rapid detection of inter-clade
229 recombination in SARS-CoV-2 with Bolotie. bioRxiv 2020; doi:
230 <https://doi.org/10.1101/2020.09.21.300913>.
- 231 [13] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
232 improvements in performance and usability. *Mol Biol Evol* 2013;30:772-780.
- 233 [14] Nguyen LT, Schmidt HA, von HA, Minh BQ. IQ-TREE: a fast and effective stochastic
234 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268-
235 274.
- 236 [15] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and
237 annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242-W245.
238
239

FIGURE LEGENDS

240

241

242 **Figure 1. Map of the nucleotide and aa substitutions and deletions of the new variant**
243 **along the SARS-CoV-2 genome (a) and three-dimensional structure of the spike protein**
244 **showing aa substitutions and deletions in various spike N501Y variants (b)**

245 a. Genes that harbor mutations are indicated by a red frame.

246 b. See supplementary Material Methods for structure prediction.

247

248 **Figure 2. Phylogenetic analysis of the SARS-CoV-2 full-length genomes of the new**
249 **variant recovered from the IHU Méditerranée Infection sequence database (doi) and**
250 **from the GISAID sequence database.**

251 The SARS-CoV-2 phylogenetic tree is based on the full-length SARS-CoV-2 genomes. The
252 SARS-CoV-2 sequences obtained by next-generation sequencing from four patients are
253 indicated by a white bold font and a dark blue background. The sequences with the highest
254 BLAST scores recovered from the NCBI GenBank nucleotide sequence databases
255 (<http://www.ncbi.nlm.nih.gov/nucleotide/>) including the 42 that are clustered with these four
256 query sequences, were incorporated in the phylogeny reconstruction, being indicated by a
257 dark blue bold font. Additional sequences included the genome of the Wuhan-Hu-1 isolate
258 and genomes obtained in our institute and classified as predominant variants. The 67 genomes
259 of SARS-CoV-2 were aligned by using MAFFT version 7 with default parameters (Nguyen et
260 al.2015). Phylogenetic tree was performed by using iQ-TREE under the GTR model and
261 visualized with iTOL [13-15].

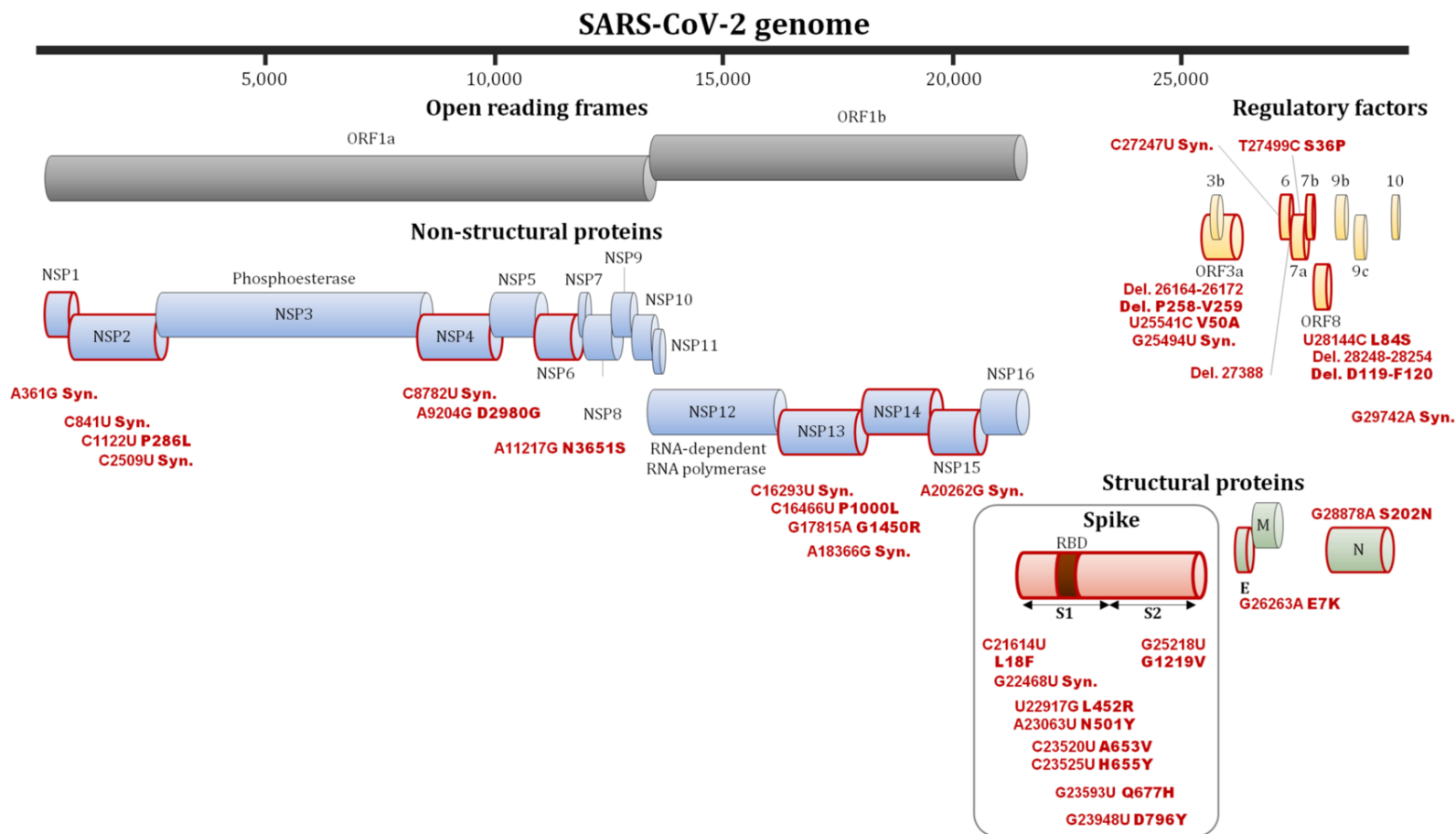
262 GISAID accession numbers: IHUCOVID-3234: EPI_ISL_1097023; IHUCOVID-3151:

263 EPI_ISL_1097024; IHUCOVID-3377: EPI_ISL_1201045; IHUCOVID-3446:

264 EPI_ISL_1201046.

Fig. 1.

a.



b.

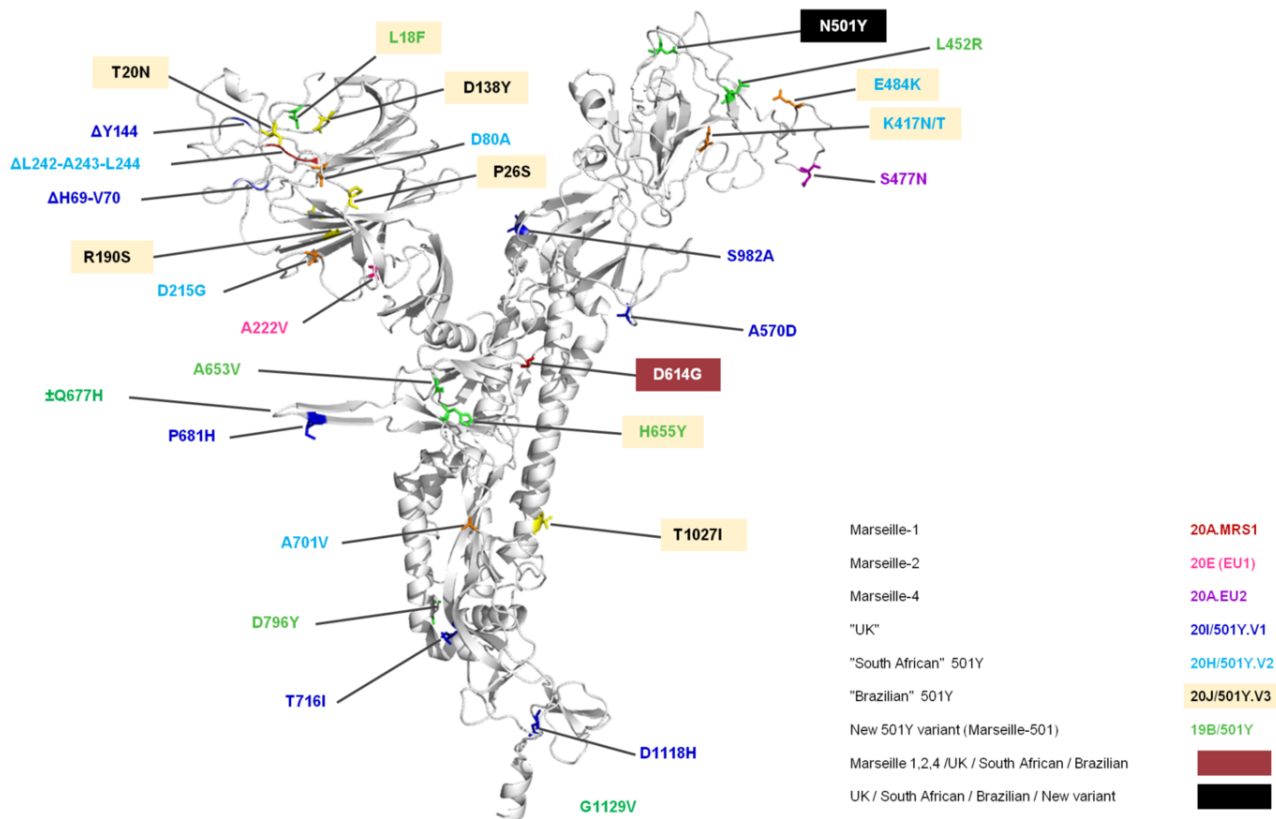


Fig. 2

Tree scale: 0.001

