

# 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

Version of Record: https://www.sciencedirect.com/science/article/pii/S1198743X21002263 Manuscript\_fab7e0743456c6a0f3bb6449a19d2c42

| 1  | TITLE PAGE  |
|----|---|
| 2  |   |
| 3  | Full-length title:  |
| 4  | Spreading of a new SARS-CoV-2 N501Y spike variant in a new lineage  |
| 5  |   |
| 6  | Short title (for the running head): SARS-CoV-2 N501Y spike variant in a new lineage   |
| 7  |   |
| 8  | Author list: Philippe COLSON <sup>1,2</sup> , Anthony LEVASSEUR <sup>1,2</sup> , Jeremy DELERCE <sup>1</sup> , Lucile           |
| 9  | PINAULT <sup>1</sup> , Pierre DUDOUET <sup>1</sup> , Christian DEVAUX <sup>1,2</sup> , Pierre-Edouard FOURNIER <sup>1,2</sup> , |
| 10 | Bernard LA SCOLA <sup>1,2</sup> , Jean-Christophe LAGIER <sup>1,2</sup> , Didier RAOULT <sup>1,2</sup> *                        |
| 11 | Affiliations: <sup>1</sup> IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille,                            |
| 12 | France; <sup>2</sup> Aix-Marseille Univ., Institut de Recherche pour le Développement (IRD),                                    |
| 13 | Assistance Publique - Hôpitaux de Marseille (AP-HM), MEPHI, 27 boulevard Jean Moulin,   |
| 14 | 13005 Marseille, France   |
| 15 |   |
| 16 | * Corresponding author: Didier Raoult, IHU - Méditerranée Infection, 19-21 boulevard  |
| 17 | Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email:                                       |
| 18 | didier.raoult@gmail.com   |
| 19 |   |
| 20 | Key words: SARS-CoV-2; variant; spike; N501Y; emergence; Covid-19; epidemic   |
| 21 |   |
| 22 | Word counts: abstract, 198; text, 1,200   |
| 23 | Figures: 2; Table: 0; References: 15  |
| 24 |   |
| 25 |   |

| 26 | ABSTRACT  |
|----|---|
| 27 |   |
| 28 | Objectives: Our surveillance of the SARS-CoV-2 genomic epidemiology led us to detect          |
| 29 | several variants since summer 2020. We report the recent spread of a new SARS-CoV-2 spike     |
| 30 | 501Y variant.   |
| 31 | Methods: SARS-CoV-2 sequences obtained from human nasopharyngeal samples by Illumina          |
| 32 | next-generation sequencing were analyzed using Nextclade and an in-house Python script and    |
| 33 | compared using BLASTn to the GISAID database. Phylogeny was performed using the IQ-           |
| 34 | TREE software.  |
| 35 | Results: We identified that SARS-CoV-2 genomes from four patients diagnosed in our            |
| 36 | institute harbored a new set of amino acid substitutions including                            |
| 37 | L18F;L452R;N501Y;A653V;H655Y;D796Y;G1219V±Q677H. These spike N501Y genomes                    |
| 38 | are the first of Nextstrain clade 19B. We obtained partial spike gene sequences of this       |
| 39 | genotype for an additional 43 patients. All patients infected with this genotype were         |
| 40 | diagnosed since mid-January 2021. We detected 42 other genomes of this genotype in            |
| 41 | GISAID, which were obtained from samples collected in December 2020 in four cases and in      |
| 42 | 2021 in 38 cases. The 89 sequences obtained in our institute or other laboratories originated |
| 43 | from the Comoros archipelago, Western European countries (mostly metropolitan France),        |
| 44 | Turkey and Nigeria.   |
| 45 | Conclusion: These findings warrant further studies to investigate the spread, epidemiological |
| 46 | and clinical features, and sensitivity to immune responses of this variant.                   |
| 47 |   |

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 1<sup>st</sup>, 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-62 pour-la-detection-par-qpcr-des-variants-sars-cov-2-marseille-4/; https://www.mediterranee-63 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 variants 20I/501Y.V1, 20H/501Y.V2, and 20J/501Y.V3, which harbor in their spike the 66 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

SARS-CoV-2 genomes were obtained directly from nasopharyngeal swab fluid by
 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 accession no. NC 045512.2 (Wuhan-Hu-1 isolate) with the CLC Genomics workbench v.7. 75 76 In addition, sequences of the spike gene 5'-region (nucleotides 1-1,854 in reference to NC\_045512.2) were obtained by NGS after PCR amplification with an in-house protocol 77 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 March 3<sup>rd</sup>, with genomes obtained in our institute as queries. Phylogeny reconstruction was 82 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 as subtitution, 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 sampled on January 18th, 2021, and its number increased from week 5 (Supplementary 115 Material Figure S1). During the January 18<sup>th</sup>-February 23<sup>rd</sup> period (37 days), SARS-CoV-2 116 117 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

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

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

131 The 47 patients we diagnosed as infected with this variant had a mean age of  $39\pm19$ 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**

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
lineages and/or to be associated with immune escape (L18F, L452R, N501Y, H655Y,
Q677H) (Supplementary Material Data). For instance, N501Y is part of several variants that

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

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

AL, JD, LP, PD, CD, PEF, BLS, JCL. Analyzed the data: PC, AL, JD, JCL, BLS, DR. Wrotethe 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

- 192
- 193

| 194 |     | REFERENCES   |
|-----|-----|--|
| 195 |     |  |
| 196 | [1] | Colson P, Levasseur A, Delerce J, Chaudet H, Bossi V, Ben Khedher M, et al. Dramatic   |
| 197 |     | increase in the SARS-CoV-2 mutation rate and low mortality rate during the second      |
| 198 |     | epidemic in summer in Marseille. IHU Preprints 2020. doi:                              |
| 199 |     | https://doi.org/10.35088/68c3-ew82.  |
| 200 | [2] | Fournier PE, Colson P, Levasseur A, Gautret P, Luciani L, Bedotto M, et al. Genome     |
| 201 |     | sequence analysis enabled deciphering the atypical evolution of COVID-19 in            |
| 202 |     | Marseille, France. medRxiv 2021. doi: https://doi.org/10.35088/kmct-tj43.              |
| 203 | [3] | Fournier PE, Colson P, Levasseur A, Devaux C, Gautret P, Bedotto M, et al. Emergence   |
| 204 |     | and outcome of the SARS-CoV-2 "Marseille-4" variant. IHU pre-prints 2021. doi:         |
| 205 |     | https://doi.org/10.35081/xcrm-6t77.  |
| 206 | [4] | Garcia-Beltran WF, Lam EC, Denis KS, Nitido AD, Garcia ZH, Hauser BM, et al.           |
| 207 |     | Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral       |
| 208 |     | immunity. medRxiv 2021. doi: https://doi.org/10.1101/2021.02.14.21251704.              |
| 209 | [5] | Colson P, Levasseur A, Gautret P, Fenollar F, Hoang VT, Delerce J et al. Introduction  |
| 210 |     | into the Marseille geographical area of a mild SARS-CoV-2 variant originating from     |
| 211 |     | sub-Saharan Africa. Travel Med Infect Dis 2021, 40, 101980.                            |
| 212 | [6] | Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: |
| 213 |     | real-time tracking of pathogen evolution. Bioinformatics 2018;34:4121-4123.            |
| 214 | [7] | Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search       |
| 215 |     | tool. J Mol Biol 1990;215:403-410.   |
| 216 | [8] | Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from      |
| 217 |     | 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. 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.

- [13] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
  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
- algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015;32:268274.

- 236 [15] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and
- annotation of phylogenetic and other trees. Nucleic Acids Res 2016;44:W242-W245.
- 238
- 239

| 240 | FIGURE LEGENDS  |
|-----|---|
| 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 NBCI 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.  |



D1118H

G1129V

Marseille 1,2,4 /UK / South African / Brazilian UK / South African / Brazilian / New variant

#### Fig. 2

