TITLE PAGE

Full-length title:

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

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

Author list: Philippe COLSON[1,2], Anthony LEVASSEUR[1,2], Jeremy DELERCE[1], Lucile PINAULT[1], Pierre DUDOUET[1], 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 (5): SARS-CoV-2; variant; spike; N501Y; L452R; emergence

Word counts: abstract, 196; text, 1,668

Figures: 3; **Tables:** 1; **References:** 30.

ABSTRACT

We report a new SARS-CoV-2 variant that is the first one classified in clade 19B to harbor a N501Y substitution in the spike protein. It has a total of seven amino acid substitutions (L18F, L452R, N501Y, A653V, H655Y, D796Y, and G1219V) including several associated with decreased sensitivity to neutralizing antibodies, and one synonymous mutation in the spike; a subgroup also harbor the Q677H substitution in the spike. We obtained a SARS-CoV-2 genome from 4 patients diagnosed in our institute, from whom the virus was cultured. In addition we obtained partial genomes covering the spike gene for 36 additional patients. All these sequences were collected since January 18th, 2021. Also, we detected 42 additional genomes currently available in the GISAID sequence database, which were obtained in 2021 in all but four cases for which they were obtained in December 2021. However, this clone has never been reported so far. These sequences originated from the Comoros archipelago, from countries in Western Europe (metropolitan France, Belgium, Denmark, England, the Netherlands), and from Turkey and Nigeria. These finding highlight the need for real-time genomic epidemiology surveillance of SARS-CoV-2 and further studies to investigate its spread and epidemiological and clinical features.

TEXT

SARS-CoV-2 has now spread worldwide since more than a year (https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases). The emergence of major viral variants expanded dramatically since summer 2020 although this only recently came into the spotlight. Since the onset of the SARS-CoV-2 outbreak in France at the end of February 2020 we have carried out epidemiological and epidemic surveillance of the genomes of these viruses by next-generation sequencing (Colson et al., 2020, Levasseur et al., 2020). Since July 2020, we have been able to detect, concomitantly with the re-appearance of SARS-CoV-2 diagnosed cases, the emergence of 10 viral variants (Fournier et al., 2021a). We focused very early on new variants as we were very skeptical on the possible occurrence of a second "wave" with the same virus. These variants have been responsible for juxtaposed or successive epidemics and have accounted for proportions of all diagnoses that varied over time. Since January 1st, 2021, we additionally performed a systematic screening by variant-specific in house qPCR assays (Bedotto et al., 2021a; Bedotto et al., 2021b), which target variants that predominate or are arising in our geographical area. These variants include the Marseille-4 variant (classified as Nexstrain clade 20A.EU1) (Fournier et al., 2021b), which has been the most prevalent between August 2020 and January 2021, and the rapidly spreading variants 20I/501Y.V1, 20H/501Y.V2, and 20J/501Y.V3, which were first detected in UK, South Africa and Brazil, respectively, and harbor in their spike the amino acid substitution N501Y (Faria, 2021).

Our genomic surveillance allowed us to recently detect SARS-CoV-2 strains harboring a new combination of eight mutations in the spike glycoprotein, based on the analysis of viral genomes recovered directly from the respiratory samples of four patients SARS-CoV-2-diagnosed in our institute, as previously described (Colson et al., 2021). These mutations

include the N501Y substitution, as well as amino acid substitutions L18F, L452R, A653V, H655Y, D796Y, and G1219V, and a synonymous mutation located at nucleotide position 22,468; in addition, a eight amino acid susbtitution in the spike, Q677H, was identified in the SARS-CoV-2 genome recovered from one patient (Figure 1). A total of 14-19 amino acid substitutions and 24-31 nucleotide substitutions are present in these strains. In addition to these full-length SARS-CoV-2 genomes, partial spike sequences with the same set of mutations were obtained from 43 other patients. We first observed this genotype from a patient sampled on January 18th, 2021, and 28 of the 47 cases were detected since February 15th. The 47 patients have a mean (±standard deviation) age of 41±22 years (range, 16-93) and 27 (57%) are male. To date,epidemiological and clinical characteristics were obtained for seven patients diagnosed in our institute as infected with this new variant (Table 1). Three travelled in or originated from Comoros, an island country in the Indian Ocean, another one originates from Guinea Conakry but did not travel abroad recently, and the three other patients did not travel and reported having been infected at work or through contact with a friend. Clinical symptoms were mild or absent. A nasopharyngeal sample from the four first case patients identified as infected with this new variant was inoculated on Vero E6 cells, as previously described (La et al., 2020), and cultures were positive in all four cases.

To assess the frequency of this genotypic pattern worldwide, we downloaded all 581,048 sequences available from the GISAID database (https://www.gisaid.org/) (Shu et al., 2017) as on March 3rd, and used them to build a database against which we performed a BLASTn search (Altschul et al., 1990) using our genome sequences as queries. We only identified 42 other genomes that share the same set of amino acid substitutions in the spike (Figure 2). These genomes were obtained from respiratory samples collected in various geographical areas. Eight of them originated from Mayotte, a French overseas department that is part of the Comoros archipelago. The other genomes originated from Denmark (n= 2),

Netherlands (n= 2), Belgium (n= 2), England (n= 1), France (n= 22; Metropolitan France, n= 21), Turkey (n= 4), and Nigeria (n= 1). Phylogenetic analysis conducted with the four genomes obtained in our institute and the 42 genomes retrieved from GISAID showed that all 46 genomes comprise a cluster delineating a 19B subclade that is strongly supported (bootstrap value, 100%) and stands apart from other clade 19B genomes (Figure 2). Overall, the delineation of this viral genotype as a new variant is supported both by the original concurrent presence of 8 mutations in the spike and by phylogenetic analysis based on genomes. In addition, genomes obtained from one patient diagnosed in our institute and from the 8 patients sampled in Mayotte comprised a cluster and all encompassed the Q677H substitution in the spike.

The seven amino acid substitutions detected alone or in combination in this new variant were present in various proportions of all sequences available on the GISAID and IHU Méditerranée Infection sequence databases. L18F, L452R, N501Y, A653V, H655Y, D796Y, and G1219V were indeed present in 36,674 (6.3%), 671 (0.1%), 17,612 (3.0%), 110 (0.02%), 1,267 (0.2%), 162 (0.03%), and 384 (0.1%), of the sequences available on GISAID on the 15th of February 2021, respectively, as determined through the CoV-GLUE online website (http://cov-glue.cvr.gla.ac.uk/#/replacement) (Chen et al., 2020). In addition, these seven substitutions were present in 14 (1.0%), 8 (0.5%), 68 (3.9%), 5 (0.29%), 5 (0.29%), 5 (0.29%), and 4 (0.23%) of the 1,744 sequences obtained in our laboratory, respectively. The four SARS-CoV-2 full-length genomes have been deposited in the GISAID database (Accession no. EPI_ISL_1097023 and EPI_ISL_1097024 have been obtained to date).

The set of spike amino acid substitutions observed in this new SARS-CoV-2 variant encompasses several substitutions reported to have recently emerged and/or to be associated with immune escape (Figure 3). The N501Y substitution that is located in the receptor binding domain of the spike protein has been reported to be associated with increased affinity

to the ACE2 cellular receptor of SARS-CoV-2 (Greaney et al., 2021). Although it appears to have alone a limited ability to affect serum antibody binding, it is part of several distinct variants detected worldwide in association with various combinations of mutations, including the three variants that rapidly spread in UK (20I/501Y.V1), South Africa (20H/501Y.V2), and Brazil (20J/501Y.V3) (Garcia-Beltran et al., 2021) and have been reported to display reduced sensitivity to neutralizing antibodies (Wang et al., 2021). The L18F substitution is located in the spike N-terminal domain that has been reported to be a target for neutralizing antibodies (Kemp et al., 2021). This substitution has been found in England in 20E (EU1) strains that expanded in Europe since summer 2020 (Hodcroft et al., 2020). In addition, it is currently rapidly spreading in a 20I/N501Y.V1 substrain with a replicative fitness reported to be increased by 1.7 fold as compared to other 20I/N501Y.V1 substrains (Grabowski et al., 2021). Also, L18F has been reported to decrease binding of neutralizing antibodies in the 20H/501Y.V2 variant (McCallum et al., 2021), and this latter variant was reported to propagate faster with than without this substitution in the presence of convalescent plasma from patients infected in South Africa in June-August, 2020 (Cele et al., 2021). Strikingly, L18F has also been found in Brazil in a majority of 20J/501Y.V3 variants obtained after November 2020 (Grabowski et al., 2021). Regarding substitution L452R, SARS-CoV-2 strains belonging to clade 20C and 20A lineages and harboring this substitution have emerged worldwide since November 2020, particularly in California (Tchesnokova et al., 2021, Zhang et al., 2021). This substitution was recently reported to reduce susceptibility to neutralizing antibodies (Li et al., 2020). The H655Y substitution was shown to confer escape from monoclonal antibodies in cell culture (Baum et al., 2020), and to arise rapidly then persist at various prevalences in experimentally-infected cats (Braun et al., 2021). The Q677H substitution recently emerged in the US in 6 distinct lineages and in clades 20G, 20A and 20B (Hodcroft et al., 2021). The D796Y substitution has been reported to occur together with the

spike double deletion ΔH69/ΔV70 following convalescent plasma therapy, and to decrease *in vitro* the susceptibility to convalescent plasma, albeit with an infectivity defect (Kemp et al., 2021). Finally, in the new variant described here, other amino acid substitutions or deletions are located in 14 other proteins than the spike; they notably include deletions in ORF3a and ORF8 genes (Figure 1).

It is currently unresolved how such strains harboring new blocks of mutations in the spike emerge. This is for instance unknown for the Marseille-4 (20A.EU2) variant or the spike N501Y-harboring variants (Fournier et al., 2021, Leung et al., 2021, Tegally et al., 2021). Overlooked genome evolution following transmission to minks and back to humans (Fournier et al., 2021b) or promotion and selection of mutations during administration of remdesivir or convalescent plasma (Kemp et al., 2021) have been suspected. Also, recombinations, which are known to be common amongst coronaviruses, may occur between SARS-CoV-2 strains (Varabyou et al., 2020). Anyway, the present variant is another example that same amino acid substitutions can occur in distinct lineages and can follow convergent evolution (Garcia-Beltran et al., 2021). This is for example the case of substitutions at spike amino acid 501 in strains of distinct lineages that emerged in distinct geographical areas, in humans as well as in minks (Garcia-Beltran et al., 2021, van Dorp et al., 2020). Epidemiologically, the countries where patients infected with the present new variant were sampled are diverse, but there is a predominant location that is the Comoros archipelago, which accounted for 11 of the 47 cases, and in France outside Mayotte. Based on limited data available here, it is unknown if this new SARS-CoV-2 variant is more or less transmissible than other strains and if it is associated with any particular clinical presentation or outcome.

Overall, these data highlight the need for genomic epidemiology surveillance of SARS-CoV-2 strains to detect new variants. The incidence of the present variant and its epidemiological and clinical features deserve being closely monitored. In addition, taking into

account its particular association of numerous spike amino acid substitutions, its sensitivity to neutralizing antibodies and to plasma from convalescent or vaccinated persons should be investigated.

Author contributions

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. Wrote the paper: PC, DR.

Funding

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

Conflicts of interest

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

Ethics

All data have been generated as part of the routine work at Assistance Publique-Hôpitaux de

Marseille (Marseille university hospitals), and this study results from routine standard clinical management. This study has been approved by our institution's ethics committee (N°2020-029). Access to the patients' biological and registry data issued from the hospital information system was approved by the data protection committee of Assistance Publique-Hôpitaux de Marseille (APHM) and was recorded in the European General Data Protection Regulation registry under number RGPD/APHM 2019-73.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **1990,** *215,* 403-410.

Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., Wei, Y., Atwal, G.S., Murphy, A.J., Stahl, N., Yancopoulos, G.D., Kyratsous, C.A. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **2020,** *369,* 1014-1018.

Bedotto, M., Fournier, P.E., Houhamdi, L., Colson, P., Raoult, D. Implementation of an in-house real-time reverse transcription-PCR assay to detect the emerging SARS-CoV-2 N501Y variants. *medRxiv* **2021;** doi: https://doi.org/10.1101/2021.02.03.21250661.

Bedotto, M., Fournier, P.E., Houhamdi, L., Levasseur, A., Delerce, J., Pinault, L., Padane, A., Chamieh, A., Tissot-Dupont, H., Brouqui, P., Sokhna, C., Azar, E., Saile, R., Mboup, S., Bitam, I., Colson, P., Raoult, D. Implementation of an in-house real-time reverse transcription-PCR assay for the rapid detection of the SARS-CoV-2 Marseille-4 variant. *medRxiv* **2021;** doi: https://doi.org/10.1101/2021.02.03.21250823

Braun, K.M., Moreno, G.K., Halfmann, P.J., Hodcroft, E.B., Baker, D.A., Boehm, E.C., Weiler, A.M., Haj, A.K., Hatta, M., Chiba, S., Maemura, T., Kawaoka, Y., Koelle, K., O'Connor, D.H., Friedrich, T.C. Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. *PLoS Pathog* **2021,** *17,* e1009373.

Cele, S., Gazy, I., Jackson, L., Hwa, S.H., Tegally, H., Lustig, G., Giandhari, J., Pillay, S., Wilkinson, E., Naidoo, Y., Karim, F., Ganga, Y., Khan, K., Balazs, A.B., Gosnell, B.I., Hanekom, W., Moosa, M.Y.,, Lessells, R.J., de Oliveira, T., Sigal, A. Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma. *medRxiv* **2021**; doi: https://doi.org/10.1101/2021.01.26.21250224.

Chen, A.T., Altschuler, K., Zhan, S.H., Chan, Y.A., Deverman, B.E. COVID-19 CG: Tracking SARS-CoV-2 mutations by locations and dates of interest. *bioRxiv* **2020**; doi: https://doi.org/10.1101/2020.09.23.310565.

Colson, P., Levasseur, A., Gautret, P., Fenollar, F., Thuan, H., V, Delerce, J., Bitam, I., Saile, R., Maaloum, M., Padane, A., Bedotto, M., Brechard, L., Bossi, V., Ben, K.M., Chaudet, H., Million, M., Tissot-Dupont, H., Lagier, J.C., Mboup, S., Fournier, P.E., Raoult, D. Introduction into the Marseille geographical area of a mild SARS-CoV-2 variant originating from sub-Saharan Africa: An investigational study. *Travel Med Infect Dis* **2021,** *40,* 101980.

Faria, N.R. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological org* **2021**; https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.

Fournier, P.E., Colson, P., Levasseur, A., Devaux, C., Gautret, P., Bedotto, M. Emergence and outcome of the SARS-CoV-2 "Marseille-4" variant. *IHU pre-prints* **2021**; doi: https://doi.org/10.35081/xcrm-6t77.

Fournier, P.E., Colson, P., Levasseur, A., Gautret, P., Luciani, L., Bedotto, M., Delerce, J., Brechard, L., Lagier, J.C., Chaudet, H., Fenollar, F., Raoult, D. Genome sequence analysis enabled deciphering the atypical evolution of COVID-19 in Marseille, France. *medRxiv* **2021**; doi: https://doi.org/10.35088/kmct-tj43.

Garcia-Beltran, W.F., Lam, E.C., Denis, K.S., Nitido, A.D., Garcia, Z.H., Hauser, B.M., Feldman, J., Pavlovic, M.N., Gregory, D.J., Poznansky, M.C., Sigal, A., Schmidt, A.G., Iafrate, A.J., Naranbhai, V., Balazs, A.B. Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *medRxiv* **2021**; doi: https://doi.org/10.1101/2021.02.14.21251704.

Grabowski, F., KochaÅ„czyk, M., Lipniacki, T. L18F substrain of SARS-CoV-2 VOC-202012/01 is rapidly spreading in England. *medRxiv* **2021** doi: https://doi.org/10.1101/2021.02.07.21251262.

Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., Bloom, J.D. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. *Cell Host & Microbe* **2021**; doi: 10.1016/j.chom.2021.02.003.

Hodcroft, E.B., Domman, D.B., Oguntuyo, K., Snyder, D.J., Diest, M.V., Densmore, K.H., Schwalm, K.C., Femling, J., Carroll, J.L., Scott, R.S., Whyte, M.M., Edwards, M.D., Hull, N.C., Kevil, C.G., Vanchiere, J.A., Lee, B., Dinwiddie, D.L., Cooper, V.S., Kamil, J.P. Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. *medRxiv* **2021**; doi: https://doi.org/10.1101/2021.02.12.21251658.

Hodcroft, E.B., Zuber, M., Nadeau, S., Comas, I., Gonzalez Candelas, F., Stadler, T., Neher, R.A. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* **2020**; Nov 27;2020.10.25.20219063. doi: 10.1101/2020.10.25.20219063.

Janson, G., Paiardini, A. PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics* **2020;** Oct 3;btaa849. doi: 10.1093/bioinformatics/btaa849. Online ahead of print.

Katoh, K., Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013,** *30,* 772-780.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **2015,** *10,* 845-858.

Kemp, S.A., Collier, D.A., Datir, R.P., Ferreira, I.A.T.M., Gayed, S., Jahun, A., Hosmillo, M., Rees-Spear, C., Mlcochova, P., Lumb, I.U., Roberts, D.J., Chandra, A., Temperton, N., Sharrocks, K., Blane, E., Modis, Y., Leigh, K., Briggs, J., van, G.M., Smith, K.G.C., Bradley, J.R., Smith, C., Doffinger, R., Ceron-Gutierrez, L., Barcenas-Morales, G., Pollock, D.D., Goldstein, R.A., Smielewska, A., Skittrall, J.P., Gouliouris, T., Goodfellow, I.G., Gkrania-Klotsas, E., Illingworth, C.J.R., McCoy, L.E., Gupta, R.K. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **2021;** Feb 5. doi: 10.1038/s41586-021-03291-y. Online ahead of print.

La, S.B., Le, B.M., Andreani, J., Hoang, V.T., Grimaldier, C., Colson, P., Gautret, P., Raoult, D. Viral RNA load as determined by cell culture as a management tool for discharge of

SARS-CoV-2 patients from infectious disease wards. *Eur J Clin Microbiol Infect Dis* **2020,** *39*, 1059-1061.

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.

Leung, K., Shum, M.H., Leung, G.M., Lam, T.T., Wu, J.T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* **2021,** *26*.

Levasseur, A., Delerce, J., Caputo, A., Brechard, L., Colson, P., Lagier, J.C., Fournier, P.E., Raoult, D. Genomic diversity and evolution of coronavirus (SARS-CoV-2) in France from 309 COVID-19-infected patients. *bioRxiv* **2020**; doi: https://doi.org/10.1101/2020.09.04.282616.

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., Qin, H., Wang, M., Lu, Q., Li, X., Sun, Q., Liu, J., Zhang, L., Li, X., Huang, W., Wang, Y. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020,** *182*, 1284-1294.

McCallum, M., Marco, A.D., Lempp, F., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., Zepeda, S., di Iulio, J., Bowen, J.E., Montiel-Ruiz, M., Zhou, J., Rosen, L.E., Bianchi, S., Guarino, B., Fregni, C.S., Abdelnabi, R., Caroline Foo, S.Y., Rothlauf, P.W., Bloyet, L.M., Benigni, F., Cameroni, E., Neyts, J., Riva, A., Snell, G., Telenti, A., Whelan, S.P.J., Virgin, H.W., Corti, D., Pizzuto, M.S., Veesler, D. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *medRxiv* **2021**; doi: https://doi.org/10.1101/2021.01.14.426475.

Nguyen, L.T., Schmidt, H.A., von, H.A., Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **2015,** *32*, 268-274.

P.Colson, A.Levasseur, J.Delerce, H.Chaudet, V.Bossi, M.Ben Khedher, P.E.Fournier, J.C.Lagier, D.Raoult**.** Dramatic increase in the SARS-CoV-2 mutation rate and low mortality rate during the second epidemic in summer in Marseille. *IHU pre-prints* **2020**; doi: https://doi.org/10.35088/68c3-ew82.

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

Tchesnokova, V., Kulakesara, H., Larson, L., Bowers, V., Rechkina, E., Kisiela, D., Sledneva, Y., Choudhury, D., Maslova, I., Deng, K., Kutumbaka, K., Geng, H., Fowler, C., Greene, D., Ralston, J., Samadpour, M., Sokurenko, E. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. *bioRxiv* **2021**; doi: https://doi.org/10.1101/2021.02.22.432189.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, M., Korsman, S., Davies, M.A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., LourenÃ§o, J., Alcantara,

L.C.J., Pond, S.L.K., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T. 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* **2021**; doi: https://doi.org/10.1101/2020.12.21.20248640.

van Dorp, L., Tan, C.C., Lam, S.D., Richard, D., Owen, C., Berchtold, D., Orengo, C., Balloux, F.o. Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv* **2020**; doi: https://doi.org/10.1101/2020.11.16.384743.

Varabyou, A., Pockrandt, C., Salzberg, S.L., Pertea, M. Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. *bioRxiv* **2020**; doi: https://doi.org/10.1101/2020.09.21.300913.

Wang, P., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., Graham, B.S., Mascola, J.R., Chang, J.Y., Yin, M.T., Sobieszczyk, M., Kyratsous, C.A., Shapiro, L., Sheng, Z., Nair, M.S., Huang, Y., Ho, D.D. Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. *bioRxiv* **2021**; doi: https://doi.org/10.1101/2021.01.25.428137.

Zhang, W., Davis, B.D., Chen, S.S., Sincuir Martinez, J.M., Plummer, J.T., Vail, E. Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA* **2021**; Feb 11;e211612. doi: 10.1001/jama.2021.1612. Online ahead of print.

FIGURE LEGEND

Figure 1. Map of the nucleotide and amino acid substitutions of the new variant along the SARS-CoV-2 genome

Genes that harbor mutations are indicated by a red frame.

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

The SARS-CoV-2 phylogenetic tree is based on the full-length SARS-CoV-2 genomes. The SARS-CoV-2 sequences obtained by next-generation sequencing from four patients are indicated by a white bold font and a dark blue background. The sequences with the highest BLAST scores recovered from the NBCI GenBank nucleotide sequence databases (http://www.ncbi.nlm.nih.gov/nucleotide/) including the 42 that are clustered with these four query sequences, were incorporated in the phylogeny reconstruction, being indicated by a dark blue bold font. Additional sequences included the genome of the Wuhan-Hu-1 isolate and genomes obtained in our institute and classified as predominant variants. The 67 genomes of SARS-CoV-2 were aligned by using MAFFT version 7 with default parameters (Nguyen et al.2015). Phylogenetic tree was performed by using iQ-TREE under the GTR model and visualized with iTOL (Katoh et al., 2013, Letunic et al., 2016, Nguyen et al., 2015).

Figure 3. Three-dimensional structure of the spike protein showing the amino acid substitutions in various variants including the new variant described here

The structure was predicted using the Phyre2 web portal (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) and visualized using the 280

372 Pymol tool v.1.8 (https://pymol.org/2/) (Janson et al., 2020; Kelley et al., 2015).

373

374

TABLES

Table 1. Virological, epidemiological and clinical features of cases of infections with the new SARS-CoV-2 variant

Strain id.	Gisaid_epi_isl	Date of sample collection	Host	Age	Gender	Country/Region of origin	Contamination setting	Covid-19 severity	Transfer in ICU	Death
IHUCOVID-3136	-	18/01/2021	Human	56	Male	Union of the Comoros	Travel abroad	Unk.	No	No
IHUCOVID-3234	EPI_ISL_1097023	21/01/2021	Human	42	Female	France	Family	Mild	No	No
IHUCOVID-3151	EPI_ISL_1097024	25/01/2021	Human	48	Female	Unk.	Unk.	Unk.	Unk.	Unk.
IHUCOVID-8134	-	02/02/2021	Human	55	Male	France	Work	Asymptomatic	No	No
IHUCOVID-3496	-	09/02/2021	Human	39	Female	France	Family	Mild	No	No
IHUCOVID-3562	-	09/02/2021	Human	16	Male	Union of the Comoros	Family	Asymptomatic	No	No
IHUCOVID-3638	-	09/02/2021	Human	25	Male	Guinea Conakry	Unk.	Mild	No	No
IHUCOVID-8230	-	15/02/2021	Human	70	Male	Union of the Comoros	Unk.	Unk.	Unk.	Unk.
IHUCOVID-1732	-	18/02/2021	Human	20	Male	Unk.	Unk.	Unk.	Unk.	Unk.

Footnote: ICU, intensive care unit; Unk., unknown