TITLE PAGE

Article type: Full-length article

Full-length title:

Emergence and outcome of the SARS-CoV-2 "Marseille-4" variant

Short title:

Outcome of the Marseille-4 genotype

Author list: Pierre-Edouard FOURNIER[1,2*], Philippe COLSON[1,3], Anthony LEVASSEUR[1,3], Christian A. DEVAUX[1,3], Philippe GAUTRET[1,2], Marielle BEDOTTO[1,3], Jeremy DELERCE[1,3], Ludivine BRECHARD[1,3], Lucile PINAULT[1,3], Jean-Christophe LAGIER[1,3], Florence FENOLLAR[1,2], Didier RAOULT[1,3]*

Affiliations: [1] IHU Méditerranée Infection; [2] Vecteurs - Infections Tropicales et Méditerranéennes (VITROME), Marseille, France ; [3] Aix-Marseille Univ, Microbes Evolution Phylogeny and Infections (MEPHI), Marseille, France.

*** Contact details for correspondence:**

Pierre-Edouard Fournier, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: pierre-edouard.fournier@univ-amu.fr

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

Keywords: SARS-CoV-2; Covid-19; variant; Marseille-4; mutations; spike; molecular epidemiology

Word count: abstract, 198; text, 2,633

Figures: 5; **Tables:** 2; **References:** 33

ABSTRACT

Introduction

In Marseille, France, following a first SARS-CoV-2 outbreak in March-May 2020, a second epidemic phase occurred from June, involving ten new variants. The Marseille-4 variant caused an epidemic that started in August and is still ongoing.

Materials and methods

The 1,038 SARS-CoV-2 whole genome sequences obtained in our laboratory by next-generation sequencing with Illumina technology were analyzed using Nextclade and nextstrain/ncov pipelines and IQ-TREE. A Marseille-4-specific qPCR assay was implemented. Demographic and clinical features were compared between patients with Marseille-4 and earlier strains.

Results

Marseille-4 harbors 13 hallmark mutations. One leads to S477N substitution in the spike receptor binding domain targeted by current vaccines. Using a specific qPCR, we observed that Marseille-4 caused 12-100% of SARS-CoV-2 infections in Marseille from September 2020, being involved in 2,106 diagnoses. This variant was more frequently associated with hypoxemia than clade 20A strains before May 2020. It caused re-infection in eleven patients SARS-CoV-2-diagnosed with different strains before June 2020, suggesting either short-term protective immunity or lack of cross-immunity.

Discussion/conclusion

Marseille-4 should be considered as a major SARS-CoV-2 variant. Its sudden appearance points toward an animal reservoir, possibly minks. The protective role of past-exposure and current vaccines against this variant should be evaluated.

TEXT

INTRODUCTION

The SARS CoV-2 epidemic that started in Wuhan, China, in December 2019, has rapidly spread around the world (https://coronavirus.jhu.edu/map.html). From January 2020, at the Mediterranee Infection institute (IHU) in Marseille, we set up the routine diagnosis of SARS-CoV-2 by PCR (Lagier et al., 2020, Colson et al., 2020a). The first SARS-CoV-2-infected patient was diagnosed at the IHU on 02/27/2020 (Colson et al., 2020c) (https://www.mediterranee-infection.com/covid-19/). Since then, we performed more than 450,000 SARS-CoV-2 PCR tests, 2,000 virus isolations by cell culture, 2,000 whole genome sequencings and took care of 14,000 positive cases. In Europe, the SARS-CoV-2 circulation was characterized by two major episodes. The first one, herein referred to as phase 1, started in February and almost ended in May (Colson et al., 2021a). However, at the end of June a second phase (phase 2) suddenly occurred, exhibiting an atypical epidemic curve which led us to suspect that the two episodes were caused by distinct viral variants. Hence, we performed whole genome sequencing of SARS-CoV-2 strains over time to characterize their genetic diversity. This enabled us to identify 10 distinct genomic patterns that successively or concomitantly spread in the Marseille area (Colson et al., 2021c; Fournier et al., 2021). Of these, two variants were identified at high frequency in the population of individuals diagnosed at the IHU. The Marseille-1 variant caused mild infections in younger patients and predominated from the end of June to the end of July 2020 (Colson et al., 2021). We accumulated evidence indicating that this variant originated in Africa and was brought to Marseille by ferry boat travelers and sailors from North Africa. In France, it did not spread outside Marseille and vanished rapidly. On July 29th, 2020, a new variant was identified and

named Marseille-4 (Figures 1, 2). We named to study the virological, clinical and epidemiological characteristics of this variant.

MATERIALS AND METHODS

Genome sequencing

Viral genomes were obtained from nasopharyngeal swab fluid using next-generation sequencing (NGS) and the Illumina Nextera XT paired-end strategy on a MiSeq instrument (Illumina Inc., San Diego, CA, USA), as previously described (Colson et al., 2021). Genome consensus sequences were assembled by mapping on the SARS-CoV-2 genome GenBank accession no. NC_045512.2 (Wuhan-Hu-1 isolate) using the CLC Genomics workbench v.7, with as thresholds 80% for nucleotide sequence coverage and 90% for nucleotide similarity. SARS-CoV-2 sequences obtained in our institute have been submitted to the GISAID database (www.gisaid.org).

Genome analysis

The 1,038 SARS-CoV-2 whole genome sequences obtained in our laboratory were analyzed using the Nextclade tool (https://clades.nextstrain.org/) (Hadfield et al., 2018) and an in-house script written in Python. Viral clades were defined on the basis of at least five available genomes sharing the same pattern of mutations. Phylogenetic trees were reconstructed by using the nextstrain/ncov tool (https://github.com/nextstrain/ncov) and visualized with the Auspice software (https://docs.nextstrain.org/projects/auspice/en/stable/). In addition, the SARS-CoV-2 genomes obtained in our laboratory were integrated in another phylogenetic analysis together with sequences from the GISAID database (www.gisaid.org) that were recovered from humans and minks. All these genomes were aligned using MAFFT v.7 (Katoh et al., 2013). Then, phylogeny reconstruction was performed using the IQ-TREE software with the GTR Model and 1,000 ultrafast bootstrap repetitions (www.iqtree.org)

(Minh et al., 2020), and the tree was visualized with the iTOL (Interactive Tree Of Life) software (https://itol.embl.de/) (Letunic et al., 2016).

PCR detection of the SARS-CoV-2 Marseille-4 variant

A qPCR system was designed that targets the nsp4 gene at nucleotide positions 9,460-9,543 in reference to genome GenBank accession no. NC_045512.2 (Wuhan-Hu-1 isolate). The primers and probe are described in Supplementary Table S1. This qPCR was run on a LC480 thermocycler (Roche Diagnostics, Mannheim, Germany). The reaction mixture contained 5 µL of 4X TaqMan Fast Virus 1-Step Master Mix (Thermo Fisher Scientific, Grand Island, NY, USA), 0.5 µL of forward primer (10 pmol/µL), 0.5 µL of reverse primer (10 pmol/µL), 0.4 µL of probe (10 pmol/µL), and 8.6 µL of water, and it was completed with 5 µL of extracted viral RNA. PCR conditions were as follows: a reverse transcription step for 10 min at 50°C, then 20 sec at 95°C followed by 40 cycles comprising a denaturation step at 95°C for 15 sec and a hybridization and elongation step at 60°C for 60 sec.

Comparisons of epidemiological and clinical features of patients diagnosed during phases 1 and 2

The demographic and clinical features of patients infected with the Marseille-4 variant were compared to those of patients infected with clade 20A strains during phase 1, between March and May 2020. Statistical analyses were carried out using R version 4.0.2. [R Core Team. R foundation for Statistical Computing, Vienna, Australia, 2020. URL: https://www.Rproject.org/].

RESULTS

Identification and circulation of the Marseille-4 variant

The highly transmissible SARS-CoV-2 Marseille-4 variant identified in Marseille at the end of July 2020 rapidly became predominant, reaching 100% of identified viral strains in

the geographical area on November 2nd. Using genome sequences available through the GISAID database (https://www.gisaid.org/), we traced back the outbreaks of this variant in different countries. The first case of infection with the Marseille-4 variant, named 20A.EU2 in the Nexstrain classification (https://clades.nextstrain.org/) (Hodcroft et al., 2020), was detected in a German patient on March 24th. Then, two cases were detected in a Balearic island, Spain, on May 29th and June 18th. Additional cases were detected in Southwestern France from July 9th, then in Denmark, and from August 1st in other European countries and other regions of France (Figures 1, 2; Supplementary Figure S1). The Marseille-4 variant was detected from September in North America (Canada, then USA), Australia and New Zealand, from October in Asia (Thailand, Hong Kong, Singapore and South Korea) and Africa (Tunisia and Morroco) and from December in Israel. In Marseille, 269 Marseille-4 complete genomes were sequenced from infected patients, and a Marseille-4-specific qPCR (Supplementary Material) was designed that enabled rapid identification of an additional 1579 cases. Overall, this variant caused 2106 cases and accounted for about two-thirds of all SARS-CoV-2 viruses tested from September 2020 to January 2021 in our place.

Genomic features

The Marseille-4 variant evolved from clade 20A strains (Figure 3) and is characterized by a combination of 20 mutations compared to the Wuhan-Hu-1 strain. Among these mutations, 13 (C4543T, G5629T, G9526T, C11497T, G13993T, G15766T, A16889G, G17019T, G22992A, C25710T, T26876C, G28975C, and G29399A) are hallmarks of this variant (Supplementary Figures S2). We provisionally subdivided the Marseille-4 variant into 11 subgroups (Marseille-4-A1 to Marseille-4-J) with a genetic drift ranging from 21 to 24 mutations compared to the Wuhan-Hu-1 strain (Table 1). Strikingly, comparative genomics shows that the set of 13 hallmark mutations appeared altogether. They are losses of a G in 7 cases and of a C in three cases, and are scattered along the viral genome. Seven (46%) are

nonsynonymous mutations, including two located in the RNA-dependent RNA polymerase (RdRp) (Nsp14; A176S and V767L), two in the NTPase/helicase (Nsp13; K1141R and E1184D), two in the nucleocapsid (N; M234I and A376T) and one in the spike glycoprotein (S; S477N). Fifteen additional mutations (C222U, C503U, G2600U, A2647G, C8937U, G18105U, C23191U, G25534U, U26442C, G26720U, G27877U, C27942U, G28086U, G29701A, G29511U) have been observed in ≥ 5 viral genomes obtained in our institute. Overall, 283 nucleotide positions are mutated in ≥ 1 Marseille-4 genomes, mostly in the Nsp3 and S genes. They were most frequently C>U (36%), G>U (25%), U>C (8%), G>A (6%), and A>G (5%) mutations, and U>- deletions (6%). Phylogenetically, the Marseille-4 variant fell within a group of viruses from Europe only (Supplementary Figures S3).

The Marseille-4 variant harbors the S477N substitution within the receptor binding domain (RBD) of the spike glycoprotein. This RBD attaches the virion to the cell membrane by binding to the viral receptor ACE2, and mediates viral entry (Lan et al., 2020). It is a major target of neutralizing antibodies (Barnes et al., 2020) and the current vaccines (Dai et al., 2020) (Figure 4). The S477N substitution has been reported to be associated with broad resistance to monoclonal neutralizing antibodies *(Liu et al.,)*. These data could explain the lack of resistance to infection by this Marseille-4 variant among people previously infected with different strains that circulated earlier, during the first phase of the 2020 pandemic. This substitution lies between substitutions observed in viruses infecting humans and others seen in viruses infecting minks (Figure 4) (Garry, 2021). It adds to the D614G substitution that was reported to increase the stability of spike trimers and confers greater affinity for ACE2 (Korber et al., 2020). It is worthy to note that the first genome available in the GISAID database (EPI_ISL_7079562020-03-24) that originates from Germany on March 24th 2020, did not harbor this S477N substitution, which may explain that it did not apparently spread further. Other critical mutations may be substitution Q57H in ORF3a, a viroporin that forms

ion channels and was reported as required for viral replication, virulence and release, and is also predicted to be a pro-apoptotic protein (Bianchi et al., 2021, Law et al., 2005), and substitutions A176S in the RdRp and K1141R and E1184D in the NTPase/helicase.

In search for the origin of the Marseille-4 variant

The origin of the Marseille-4 variant is currently unknown. It emerged abruptly with its block of specific mutations, with no known intermediate form, while the SARS-CoV-2 epidemic had almost ended in France and Europe (Figure 1, Figure 3). This apparently discontinuous evolution of SARS-CoV-2 genomes is abnormal, particularly if we consider that after its first detection this variant had shown a subsequent mutation rate similar to that of other lineages (e.g., mutation in the RdRp did not alter the polymerase fidelity). Although we cannot exclude that the missing intermediate exists but has not been sequenced so far from COVID-19 patients, this could also suggest that there is an overlooked reservoir in which the virus was submitted to a selection pressure that favored a particular increase in mutation accumulation. Interestingly, among the 10,516 sequences from the Marseille-4 variant in the GISAID database (on January 24th, 2021), the 272 genomes from our laboratory had close relatives with those originating from Northern Europe, mostly Denmark (3,366), the UK (2,652) and Switzerland (1,147) (Supplementary Figure S1). A phylogenetic tree was constructed that included genomes from mink and human SARS-CoV-2 strains. Mink strains were divided into five and six main groups, for the samples from the Netherlands and Denmark, respectively (Figure 5). We observed a common phylogenetic node between mink strains, the Marseille-4, Marseille-5, Marseille-6 variants and the 20H/501Y.V2 variant from England. This node pointed to the above-described common mutation, Q57H in ORF3a. The rapid emergence of the Marseille-4 variant during summer 2020, after the end of the first epidemic phase, may point toward an animal reservoir. Mink farms were identified as reservoirs and sources of SARS-CoV-2 mutants in the Netherlands in April (Oude Munnink et

al., 2021), and in Denmark in June 2020 (Hammer et al., 2020). In France, one of the four mink farms was infected and animals were culled. SARS-CoV-2 is an epizootic agent that caused an outbreak in humans before being transferred to mink in which it spread rapidly through densely caged animals and subsequently became a source for human infection. To date, more than 800 human infections from minks have been reported (Oude Munnink et al., 2021). One hypothesis could be that a human SARS-CoV-2 from infected caregivers infected mink, then the frequency of viral mutations changed in the mink due to a different host selection pressure, and this mink-adapted virus (with multiple mutations) became a new viral source to infect humans. The genome obtained from a German patient sampled on March 24th (EPI_ISL_7079562020-03-24) is atypical as it is devoid of the S477N substitution, one of the Marseille-4 hallmark mutations, but harbors more mutations (n= 31) than the other Marseille-4 strains, including in Nsp2, Nsp3, S and N proteins, and in ORF1b, particularly the Nsp14 exonuclease, which has a proofreading activity (Shannon et al., 2020). The evolutionary relationships of this genome with other Marseille-4 genomes warrants a further investigation with the availability of other genomes obtained from samples collected during the same period.

Clinical findings: Marseille-4 variant may escape immunity conferred by a first SARS-CoV-2 infection

Compared to the clade 20A strains that predominated during phase 1 between March and May 2020, the Marseille-4 variant was associated with a lower frequency of cough, rhinitis and olfactory and gustatory disorders (Table 2). By contrast, hypoxemia was more frequent in patients infected with Marseille-4 variant. It was reported that differences observed in COVID-19 severity may in part be associated with the dysfunction of cellular immune responses to SARS-CoV-2 and/or a weakness of neutralizing humoral response (Moderbacher et al., 2020). We diagnosed two successive COVID-19 infections, separated by

more than 4 months, in 11 patients. The first infection was diagnosed before June 2020 when Marseille-4 was circulating in Marseille (Colson et al., 2020b; Brouqui et al., 2021), and we obtained genomic or qPCR (1 and 10 patients, respectively) confirmation that the second episode was caused by the Marseille-4 variant. This suggests either a short protective immunity (only a few weeks or months) as previously observed with seasonal coronaviruses (Edridge et al., 2020), or a lack of cross-immunity between different SARS-CoV-2 variants, allowing Marseille-4 to evade immune protection elicited by another earlier variant. This may be related to the S477N mutation which could change the affinity of RBD for ACE2 and decrease the sensitivity of the variant virus to anti-RBD-specific neutralizing antibodies (Andreano et al., 2020).

DISCUSSION

The recent evolution of the SARS-CoV-2 epidemics reflects the generation of new variants in different ecosystems that spread with globalization and replaced the original variants issued from Wuhan. Some can be associated with different clinical features as for the case of the Marseille-4 variant. The ecosystems allowing this selection may consist of human groups isolated for a while, or animal reservoirs such as minks in large farms. Large concentrations of farmed minks were infected by human SARS-CoV-2 (Oude Munnink et al., 2020). Under these conditions, sub-speciation may occur (Darwin, 1859). The re-connection of isolated ecosystems (either countries and/or farmed animals) where different variants had developed generated new outbreaks in countries that were exposed to incoming populations such as travelers. Several reasons lead us to believe that minks were the source of the Marseille-4 variant. First, this variant carries a new set of several mutations which seems to have appeared suddenly based on the analysis of all the genomes available worldwide, and not gradually. This suggests that this brutal genome evolution had been overlooked. Secondly,

there was no SARS-CoV-2 epidemic in France at the time of the emergence of this variant, except in a region near the city of Laval (Mayenne, Western France) located between the most dense area in wild minks (Brittany) and a mink farm (Eure-et-Loire) where 30% of minks were proved to be SARS-CoV-2-positive by qPCR and 97% had antibodies against the virus. As a consequence, the entire farm mink population was slaughtered (https://www.plateforme-esa.fr/article/covid-19-et-animaux-mise-a-jour-au-05-01-2021; Fenollar et al., 2021). Progressively, this SARS-CoV-2 epidemic spread in France during the summer, and we observed the first cases of Marseille-4 infections in Marseille when French tourists arrived in our region. For unknown reasons, the sequence of the virus of the farm minks infected mid-November is not yet available.

CONCLUSION

Overall, we believe that the segregation of viral strains in isolated geographical areas and in animal reservoirs may contribute to explain the differences observed among epidemic curves around the world. This would help to understand the mechanism of the second episode that developed in Marseille, initially caused by an African variant that disappeared (Colson et al., 2021), and then by emerging new variants linked to different areas of Europe, including those hosting huge mink farms. Finally, the role of the treatment of COVID-19 by remdesivir or hyperimmune plasma (Choi et al., 2020, Kemp et al., 2020) in generating and selecting variants may also have contributed to the new outbreaks observed in the most developed countries.

AUTHORSHIP CONTRIBUTION STATEMENT

Conceived and designed the experiments: DR, PEF, PC and PG. Contributed materials/analysis tools: PEF, PC, AL, CD, PG, MB, JD, LB, LP, JCL, FF. Analyzed the data:

PEF, PC, AL, PG, JD, JCL, FF, DR. Wrote the paper: PEF, PC, CD, PG, DR. All authors approved the last version of the manuscript.

Acknowledgments

We are grateful to Olivia Ardizzoni, Vincent Bossi, Madeleine Carrera, Vera Esteves-Vieira, Laurence Thomas, Priscilla Jardot and Raphael Tola for their technical help, and to Audrey Giraud-Gatineau and Léa Luciani for their help with data analysis. This manuscript The manuscript text has been edited by a native English speaker.

Ethical approval

The study was approved by the ethical committee of the Méditerranée Infection institute under references No. 2020-016-3. 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.

Funding

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

Competing interests

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.

Data and materials availability

Data underlying the study are available from the GISAID database (https://www.gisaid.org/) or from the corresponding author upon request.

REFERENCES

Andreano E, Piccini G, Licastro D, Johnson NV, Paciello I, Monego SD, et al. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. bioRxiv 2020; doi: https://doi.org/10.1101/2020.12.28.424451.

Barnes CO, West AP, Jr., Huey-Tubman KE, Hoffmann MAG, Sharaf NG, Hoffman PR et al. Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. Cell 2020;182:828-842.

Bianchi M, Borsetti A, Ciccozzi M, Pascarella S. SARS-Cov-2 ORF3a: Mutability and function. Int J Biol Macromol 2021;170:820-826.

Brouqui P, Colson P, Melenotte C, Houhamdi L, Bedotto M, Devaux C, et al. COVID-19 re-infection. Eur J Clin Invest 2021 Mar 6:e13537. doi: 10.1111/eci.13537. Epub ahead of print.

Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF, Qiu X, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised Host. N Engl J Med 2020;383:2291-2293.

Colson P, Esteves-Vieira V, Giraud-Gatineau A, Zandotti C, Filosa V, Chaudet H, et al. Temporal and age distributions of SARS-CoV-2 and other coronaviruses, southeastern France. Int J Infect Dis 2020a;101:121-125.

Colson P, Finaud M, Levy N, Lagier JC, Raoult D. Evidence of SARS-CoV-2 re-infection with a different genotype. J Infect 2020b Nov 15;S0163-4453(20)30706-4. doi: 10.1016/j.jinf.2020.11.011. Online ahead of print.

Colson P, Lagier JC, Baudoin JP, Bou KJ, La Scola B, Raoult D. Ultrarapid diagnosis, microscope imaging, genome sequencing, and culture isolation of SARS-CoV-2. Eur J Clin Microbiol Infect Dis 2020c;39:1601-1603.

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 pre-prints 2020c; doi: https://doi.org/10.35088/68c3-ew82.

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.

Dai L, Gao GF. Viral targets for vaccines against COVID-19. Nat Rev Immunol 2020;1-10.

Darwin, C. 1859. On the origin of species. John Murray, London.

Edridge AWD, Kaczorowska J, Hoste ACR, Bakker M, Klein M, Loens K, et al. Seasonal coronavirus protective immunity is short-lasting. Nat Med 2020;26:1691-1693.

Fenollar F, Mediannikov OY, Maurin M, Devaux CA, Colson P, Levasseur A, et al. Mink, SARS-CoV-2, and the human-animal interface. Front Microbiol 2021, in press.

Fournier PE, Colson P, Levasseur A, Gautret P, Bedotto M, Filosa V, et al. Genome sequence analysis enabled deciphering the atypical evolution of COVID-19 epidemics in Marseille, France. IHU pre-prints 2021; doi: https://doi.org/10.35088/kmct-tj43.

Garry RF. Mutations arising in SARS-CoV-2 spike on sustained human-to-human transmission and human-to-animal passage. Virological org 2021; https://virological.org/t/mutations-arising-in-sars-cov-2-spike-on-sustained-human-to-human-transmission-and-human-to-animal-passage/578.

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.

Hammer AS, Quaade ML, Rasmussen TB, Fonager J, Rasmussen M, Mundbjerg K et al. SARS-CoV-2 Transmission between Mink (*Neovison vison*) and Humans, Denmark. Emerg Infect Dis 2021;27:547-551. doi: 10.3201/eid2702.203794. Epub 2020 Nov 18.

Hodcroft EB, Zuber M, Nadeau S, Comas I, Gonzalez Candelas F, SeqCOVID-SPAIN consortium, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv 2020;https://doi.org/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 DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772-780.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 2015;10:845-858.

Kemp SA, Collier DA, Datir R, Ferreira I, Gayed S, Jahun A, et al. Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. medRxiv 2020; doi: 10.1101/2020.12.05.20241927.

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 2020;182:812-827.

Lagier JC, Million M, Gautret P, Colson P, Cortaredona S, Giraud-Gatineau A, et al. Outcomes of 3,737 COVID-19 patients treated with hydroxychloroquine/azithromycin and other regimens in Marseille, France: A retrospective analysis. Travel Med Infect Dis 2020;36:101791.

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 2020;581:215-220.

Law PTW, Wong CH, Au TCC, Chuck CP, Kong SK, Chan PKS, et al. The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells. J Gen Virol 2005;86:1921-1930.

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.

Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, et al. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. bioRxiv 2020; doi: 10.1101/2020.11.06.372037

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von HA, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 2020;37:1530-1534.

Moderbacher CR, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, et al. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. Cell 2020;183:996-1012.

Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science 2021;371:172-177.

Shannon A, Le NT, Selisko B, Eydoux C, Alvarez K, Guillemot JC, et al. Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. Antiviral Res 2020;178:104793. doi: 10.1016/j.antiviral.2020.104793.:104793.

FIGURE LEGENDS

Figure 1. Schematic of the evolution of the SARS-CoV-2 Marseille-4 variant in Europe.

Figure 2. Evolution of the Marseille-4 variant over time.

a. Weekly number of genomes of the Marseille-4 variant worldwide

b. Weekly frequency normalized to 100% of the countries where genomes of the Marseille-4 variant were obtained

c. Time distribution of the daily number of genomes of the Marseille-4 variant per country

d. Weekly number of genomes of the Marseille-4 variant in French regions

e. Weekly frequency normalized to 100% of the French regions where genomes of the Marseille-4 variant were obtained

Figure 3. Genome sequence-based phylogenetic trees showing the evolution of SARS-CoV-2 Marseille-4 variant strains.

Full-length genome sequences obtained in our study were compared to those available in the GISAID database (https://www.gisaid.org/). Phylogenetic trees were reconstructed and visualized by using the Nextstrain pipeline (https://github.com/nextstrain/ncov/) (Hadfield et al., 2018). a. Time-scale phylogenetic tree. b. Phylogenetic tree based on mutational events.

Figure 4. 3D structure of the spike protein showing the amino acid substitutions in the receptor-binding motif of the Marseille-4 variant and of other variants detected in humans and/or minks

The structure was predicted using the Phyre2 web portal (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) (Kelley et al., 2015) and

visualized using the Pymol tool v.1.8 (https://pymol.org/2/) (Janson et al., 2020). Amino acids where a substitution was observed in humans are colored in red, where a substitution was observed in minks are colored in yellow, and those where a substitution was observed in humans and minks are colored in orange.

Figure 5. Phylogenetic tree based on SARS-CoV-2 full-length genomes.

A total of 744 genomes of SARS-CoV2 were integrated in a phylogenetic analysis. All genomes were aligned using MAFFT version 7 (Katoh et al., 2013). Phylogenetic tree was reconstructed by using IQ-TREE with the GTR model with 1,000 ultrafast bootstrap repetitions (Minh et al., 2020), and visualized with iTOL (Interactive Tree Of Life, (https://itol.embl.de/)) (Letunic et al., 2016).

DK, Denmark; NTH, The Netherlands.

TABLES

Table 1. Nucleotide mutations and amino acid substitutions in the genomes of SARS-CoV-2 Marseille-4 variants

Nucleotide position	Gene	WT nt	Mutated nt	Codon change	Codon number	Amino acid substitution	Wuhan	20A	20A/25563T	20A/18877T	20A/26735T	Marseille4	Marseille4-A	Marseille4-A1	Marseille4-B	Marseille4-C	Marseille4-D	Marseille4-E	Marseille4-F	Marseille4-G	Marseille4-H	Marseille4-I	Marseille4-J
												IHUCOVID-1019	IHUCOVID-1164	IHUCOVID-1363	IHUCOVID-1588	IHUCOVID-2056	IHUCOVID-1388	IHUCOVID-2393	IHUCOVID-1377	IHUCOVID-1569	IHUCOVID-1630	IHUCOVID-1908	IHUCOVID-2295
								0	4	5	6	7	20	21	22	24	21	21	22	21	21	21	21	22
222	5UTR	C	T				17	1	C	C	C	C	C	C	C	T	C	C	C	C	C	C	N	C
241	5UTR	C	T				1016	11	C	T	T	T	T	T	T	T	T	T	T	T	T	T	N	T
503	nsp1	C	T	CCT>TCT	80	P80S	17	1	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C
2600	nsp2	G	T	GTT>TTT	599	V599F	9	1	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G
2647	nsp2	A	G	AAA>AAG	614	K	7	1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G
3037	nsp3	C	T	TTC>TTT	106	F	1030	12	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
4543	nsp3	C	T	ACC>ACT	608	T	268	12	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T
5629	nsp3	G	T	ACG>ACT	970	T	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T
6539	nsp3	C	T	CAC>TAC	1274	H1274Y	6	1	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C
8937	nsp4	C	T	GCA>GTA	128	A128V	4	1	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C
9526	nsp4	G	T	ATG>ATT	324	M324I	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T
11497	nsp6	C	T	TAC>TAT	175	Y	268	12	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T
13993	nsp12b	G	T	GCT>TCT	176	A176S	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T
14408	nsp12b	C	T	CCT>CTT	314	P314L	1028	12	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
15766	nsp12b	G	T	GTG>TTG	767	V767L	268	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T
16889	nsp13	A	G	AAA>AGA	218	K218R	269	12	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G
17019	nsp13	G	T	GAG>GAT	261	E261D	269	12	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T
18105	nsp14	G	T	CAG>CAT	22	Q22H	6	1	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G
18877	nsp14	C	T	CTA>TTA	280	L	272	12	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T
22992	S	G	A	AGC>AAC	477	S477N	269	12	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A
23191	S	C	T	TTC>TTT	543	F	13	1	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C
23403	S	A	G	GAT>GGT	614	D614G	1028	12	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
25534	ORF3a	G	T	GTT>TTT	48	V48F	7	1	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T
25563	ORF3a	G	T	CAG>CAT	57	Q57H	657	12	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T
25710	ORF3a	C	T	CTC>CTT	106	L	272	12	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T
26442	E	T	C	AAT>AAC	66	N	6	1	T	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T
26720	M	G	T	GTG>GTT	66	V	16	1	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G
26735	M	C	T	TAC>TAT	71	Y	272	12	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
26876	M	T	C	ATT>ATC	118	I	269	12	T	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C
27877	ORF7b	G	T	TGT>TTT	41	C41F	6	1	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G
27942	ORF8	C	T	CAC>TAC	17	H17Y	12	1	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C
28086	ORF8	G	T	GCT>TCT	65	A65S	30	2	G	G	G	G	G	G	T	T	G	G	G	G	G	G	G	G
28975	N	G	C	ATG>ATC	234	M234I	268	12	G	G	G	G	G	C	C	C	C	C	C	C	C	C	C	C
29399	N	G	A	GCT>ACT	376	A376T	263	12	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A
29511	N	G	T	AGT>ATT	413	S413I	6	1	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G
29701	3UTR	G	A				12	1	G	G	G	G	G	G	G	G	G	N	G	G	A	G	G	G

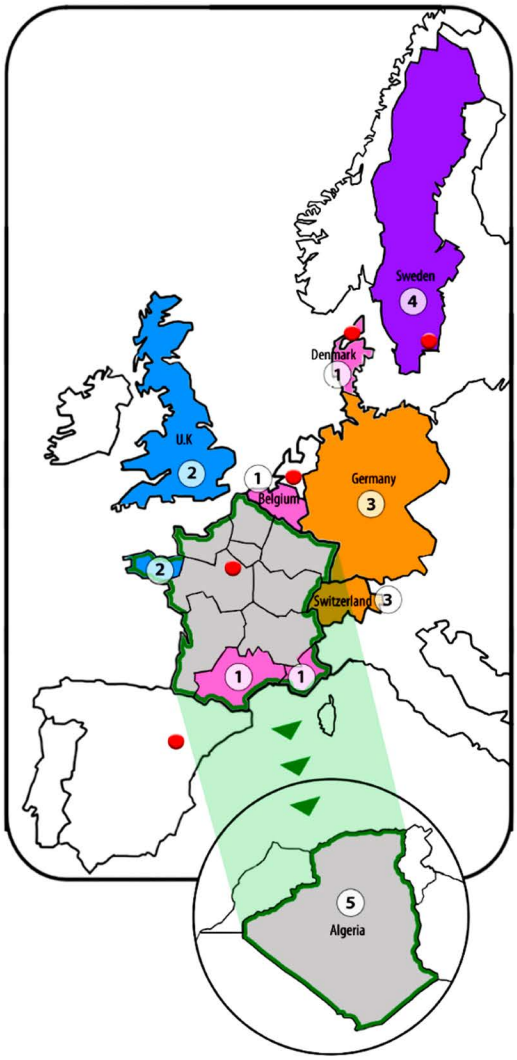
Table 2. Demographics, outcomes and clinical symptoms in patients infected with different SARS-CoV-2 variants

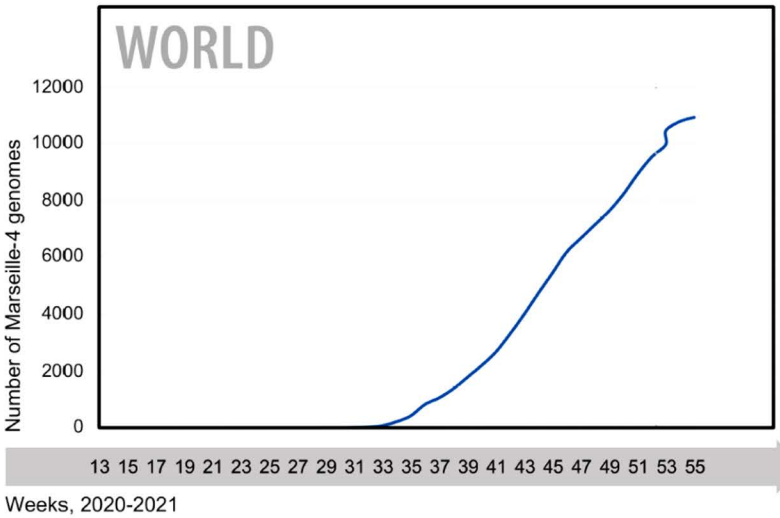
Demographics and outcomes (N=759)	20A (N=339)		Marseille-4 (N=420)		p-value*
	n	%	n	%	
Male gender	151	44.5	216	51.4	0.059
Age (mean ± SD)	50.2 ± 22.3		48.9 ± 23.1		0.41
Hospitalization	53	15.6	68	16.2	0.835
Transfer to intensive care unit	5	1.5	10	2.4	0.44
Death	10	2.9	16	3.8	0.52
Symptoms (N=444)	**20A (N=254)**		**Marseille-4 (N=190)**		**p-value***
	n	%	n	%	
Cough	123	48.4	73	38.4	**0.036**
Rhinitis	106	41.7	37	19.5	**<0.0001**
Anosmia	76	29.9	35	18.5	**0.006**
Ageusia	71	27.9	34	18.0	**0.015**
Dyspnea	72	28.3	42	22.1	0.136
SpO2 <96%	37	14.6	42	22.1	**0.04**

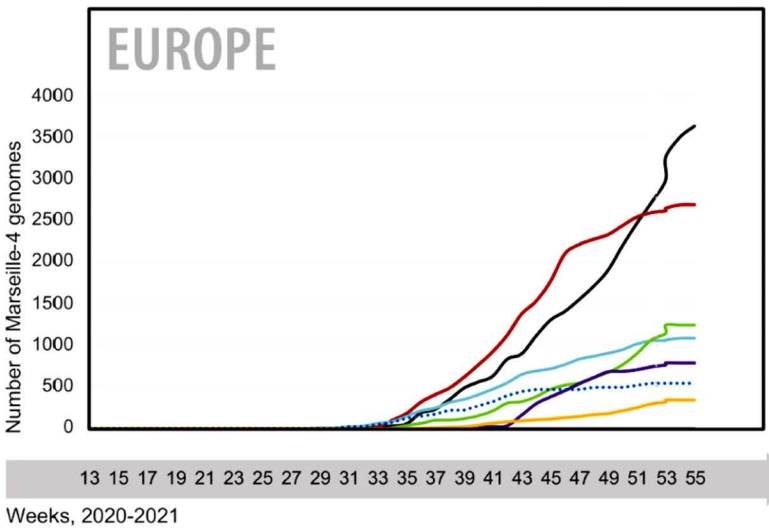
* Chi2 or Fisher exact test for qualitative variables. Student test for quantitative variables

Fig.1

Epidemic of the Marseille-4 variant







1 Denmark, South France, Belgium : Week 28 - Week 30

2 Bretagne France, UK : Week 30 - Week 32

3 Switzerland, Germany : Week 33 - Week 34

4 Sweden : Week 34 - Week 35

5 Algeria : Week 38

Mink farming

IHUMI_Marseille

Denmark

Switzerland

UK

Luxembourg

France

Netherlands

Fig. 2

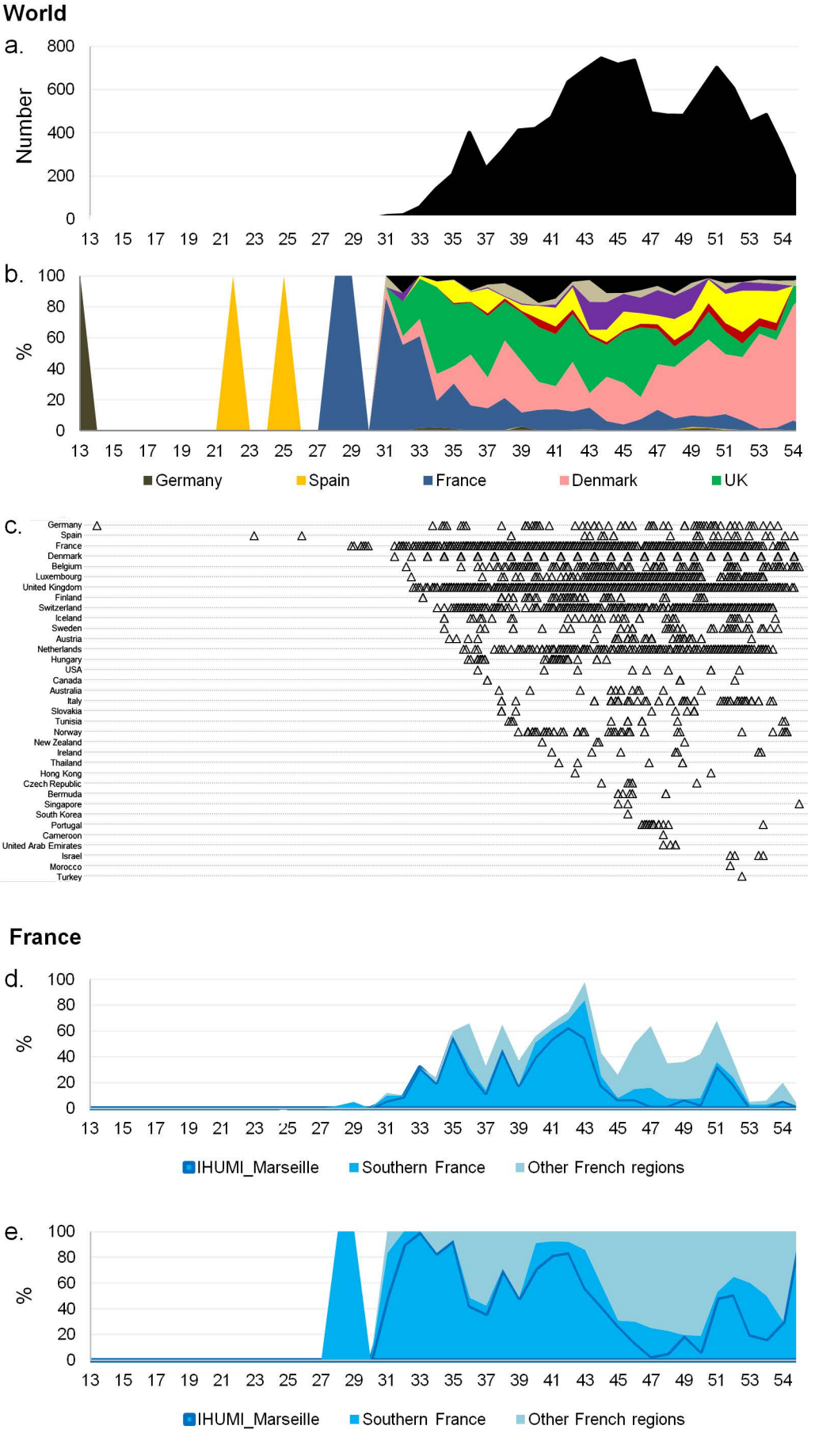
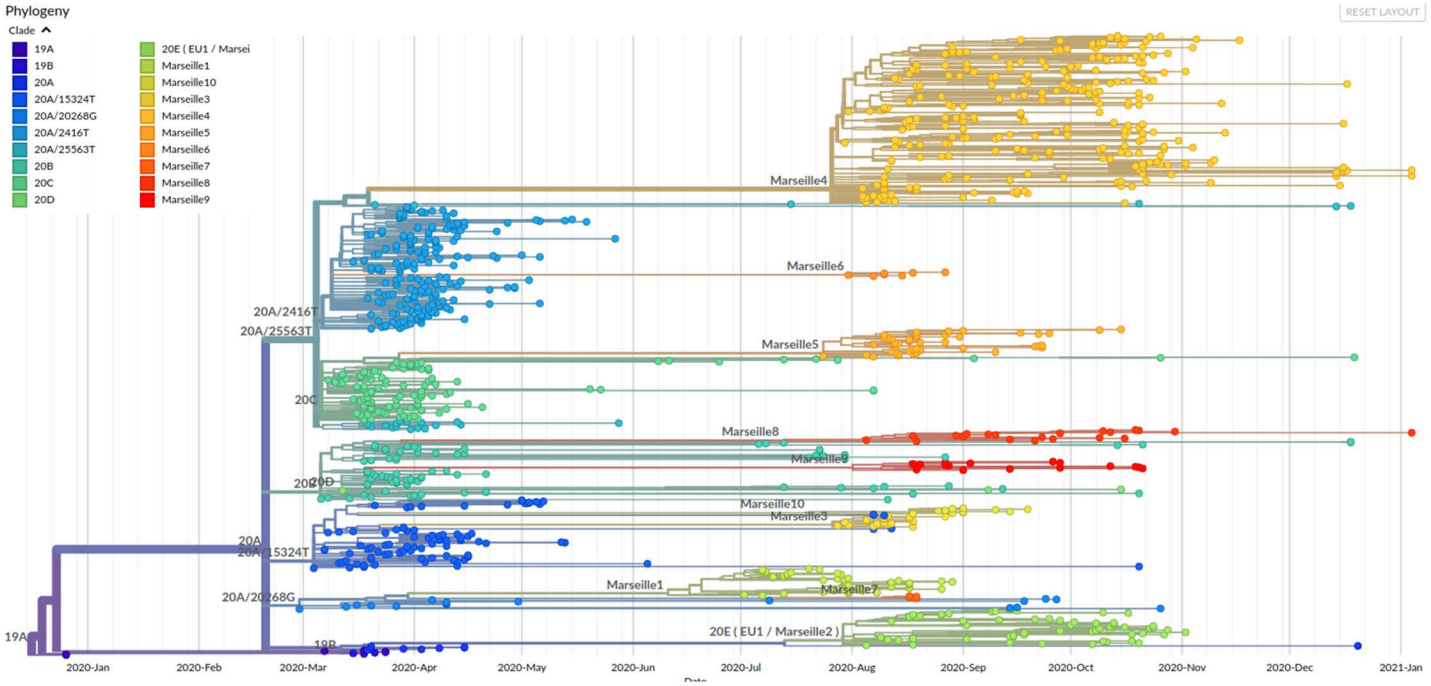

World
a.
Number
800
600
400
200
0
13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 54
b.
%
100
80
60
40
20
0
13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 54
Germany
Spain
France
Denmark
UK
c.
Germany
Spain
France
Denmark
Belgium
Luxembourg
United Kingdom
Finland
Switzerland
Iceland
Sweden
Austria
Netherlands
Hungary
USA
Canada
Australia
Italy
Slovakia
Tunisia
Norway
New Zealand
Ireland
Thailand
Hong Kong
Czech Republic
Bermuda
Singapore
South Korea
Portugal
Cameroon
United Arab Emirates
Israel
Morocco
Turkey
France
d.
%
100
80
60
40
20
0
13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 54
IHUMI_Marseille
Southern France
Other French regions
e.
%
100
80
60
40
20
0
13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 54
IHUMI_Marseille
Southern France
Other French regions

Fig. 3

a.


Phylogeny
Clade
19A
19B
20A
20A/15324T
20A/20268G
20A/2416T
20A/25563T
20B
20C
20D
20E (EU1 / Marsei
Marseille1
Marseille10
Marseille3
Marseille4
Marseille5
Marseille6
Marseille7
Marseille8
Marseille9
RESET LAYOUT
Marseille4
Marseille6
20A/2416T
20A/25563T
Marseille5
20C
Marseille8
Marseille9
20D
Marseille10
Marseille3
20A
20A/15324T
Marseille1
Marseille7
20A/20268G
20E (EU1 / Marseille2)
19A
19B
2020-Jan
2020-Feb
2020-Mar
2020-Apr
2020-May
2020-Jun
2020-Jul
2020-Aug
2020-Sep
2020-Oct
2020-Nov
2020-Dec
2021-Jan

b.

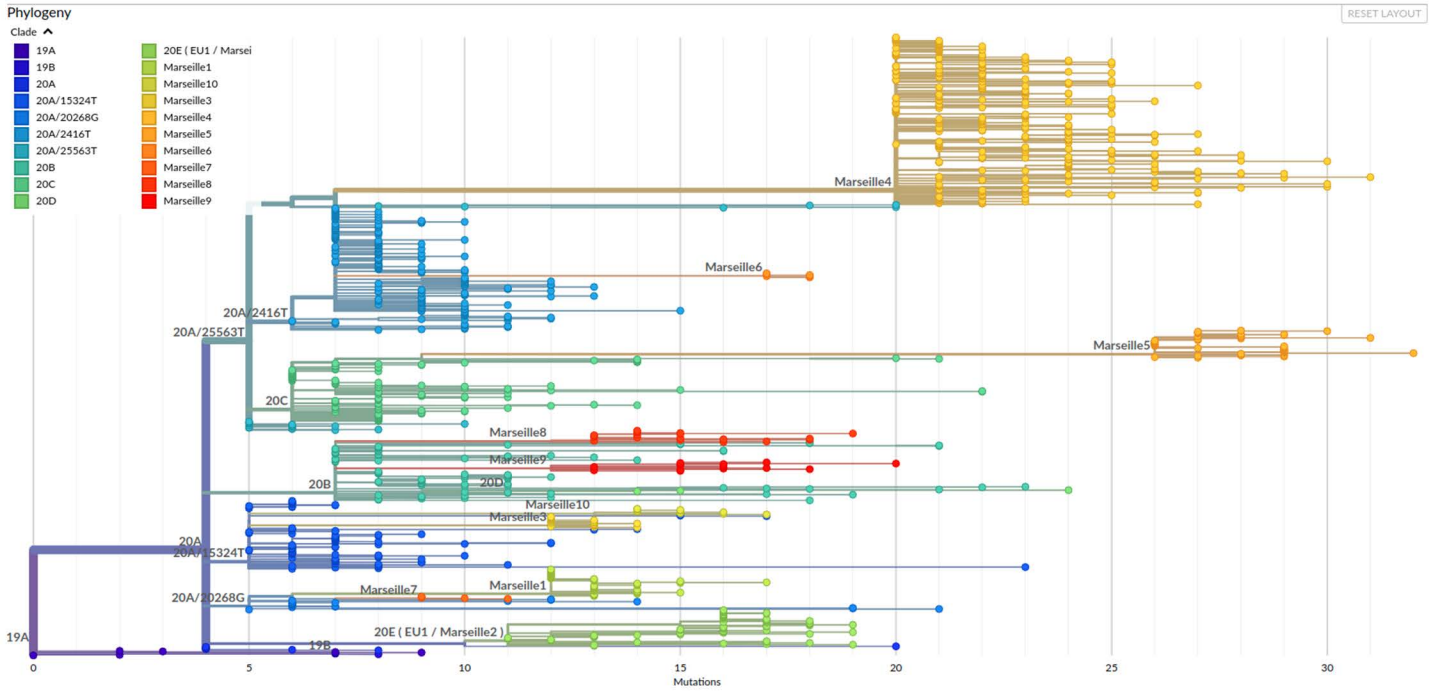

Phylogeny
Clade
19A
19B
20A
20A/15324T
20A/20268G
20A/2416T
20A/25563T
20B
20C
20D
20E (EU1 / Marsei
Marseille1
Marseille10
Marseille3
Marseille4
Marseille5
Marseille6
Marseille7
Marseille8
Marseille9
RESET LAYOUT
Marseille4
Marseille6
20A/2416T
20A/25563T
Marseille5
20C
Marseille8
Marseille9
20B
20D
Marseille10
Marseille3
20A
20A/15324T
Marseille7
Marseille1
20A/20268G
20E (EU1 / Marseille2)
19A
19B
0
5
10
15
20
25
30
Mutations

Fig. 4

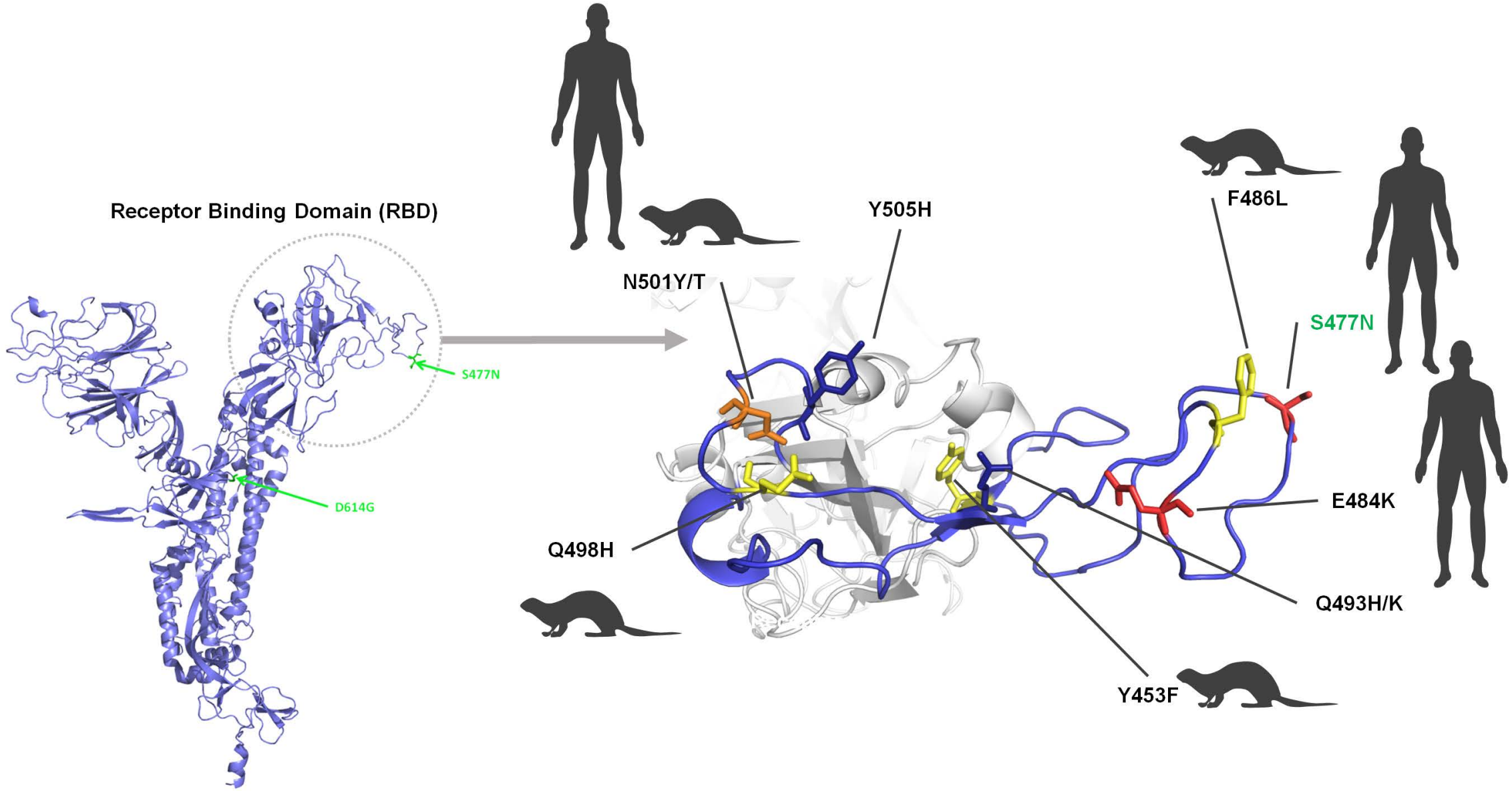
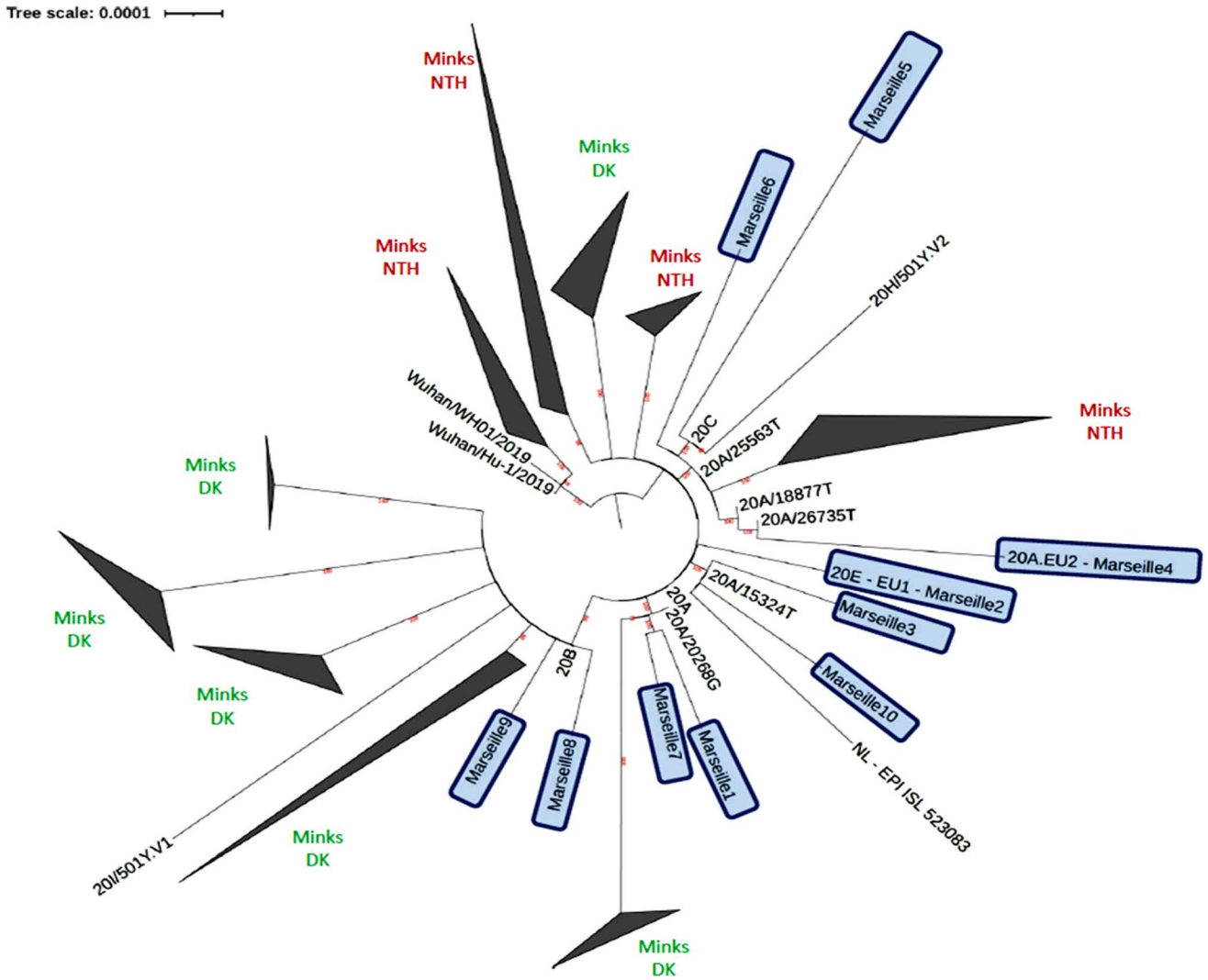

Receptor Binding Domain (RBD)
S477N
D614G
N501Y/T
Y505H
F486L
S477N
E484K
Q498H
Q493H/K
Y453F

Fig. 5


Tree scale: 0.0001
Minks NTH
Minks DK
Minks NTH
Minks NTH
Marseille5
Marseille6
20H/501Y.V2
Minks NTH
Wuhan/WH01/2019
Wuhan/Hu-1/2019
20C
20A/25563T
20A/18877T
20A/26735T
20A.EU2 - Marseille4
20E - EU1 - Marseille2
Marseille3
20A/15324T
20A
20A/20268G
Marseille10
NL - EPI ISL 523083
Minks DK
Minks DK
Minks DK
Minks DK
20B
Marseille9
Marseille8
Marseille7
Marseille1
20I/501Y.V1
Minks DK
Minks DK

TITLE PAGE FOR THE SUPPLEMENTARY MATERIAL

Article type: Full-length article

Full-length title:

Emergence and outcome of the SARS-CoV-2 "Marseille-4" variant

Short title: Outcome of the Marseille-4 genotype

Author list: Pierre-Edouard FOURNIER[1,2*], Philippe COLSON[1,3], Anthony LEVASSEUR[1,3], Christian A. DEVAUX[1,3], Philippe GAUTRET[1,2], Marielle BEDOTTO[1,3], Jeremy DELERCE[1,3], Ludivine BRECHARD[1,3], Lucile PINAULT[1,3], Jean-Christophe LAGIER[1,3], Florence FENOLLAR[1,2], Didier RAOULT[1,3]*

Affiliations: [1] IHU Méditerranée Infection; [2] Vecteurs - Infections Tropicales et Méditerranéennes (VITROME), Marseille, France ; [3] Aix-Marseille Univ, Microbes Evolution Phylogeny and Infections (MEPHI), Marseille, France.

*** Contact details for correspondence:**

Pierre-Edouard Fournier, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: pierre-edouard.fournier@univ-amu.fr

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

SUPPLEMENTARY MATERIAL

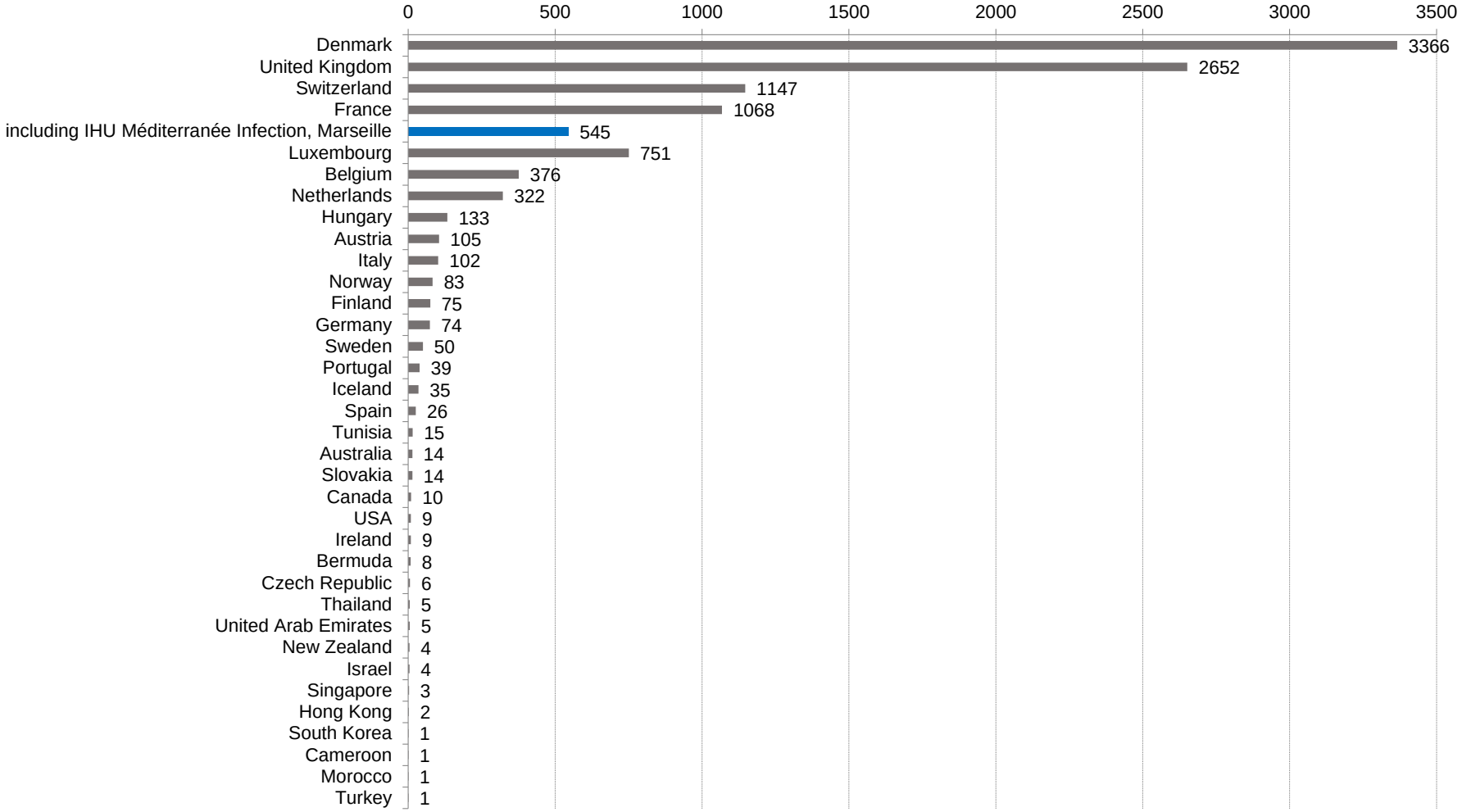
Supplementary Figures: 3

Supplementary Tables: 1

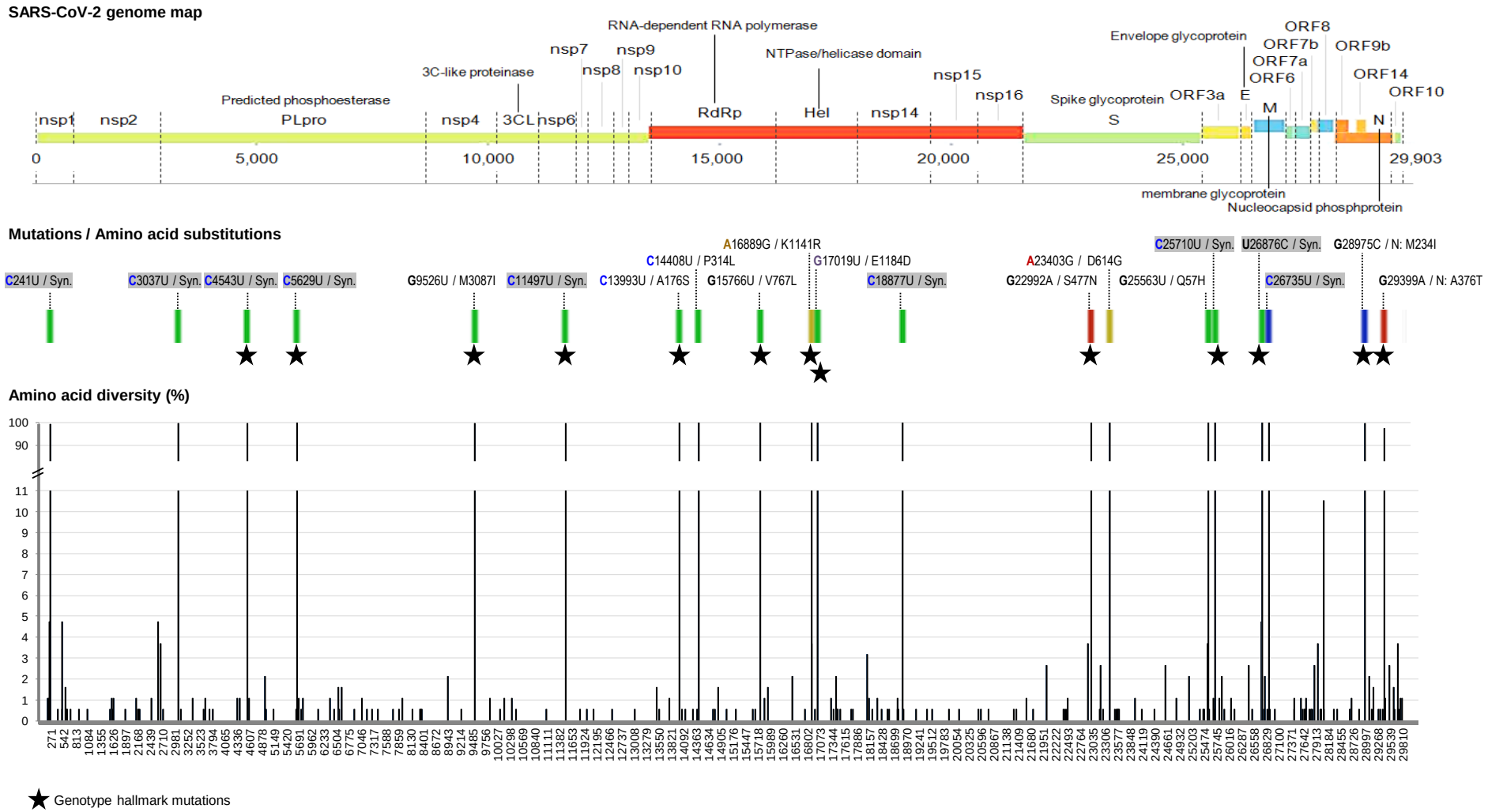
References: 2

SUPPLEMENTARY FIGURES

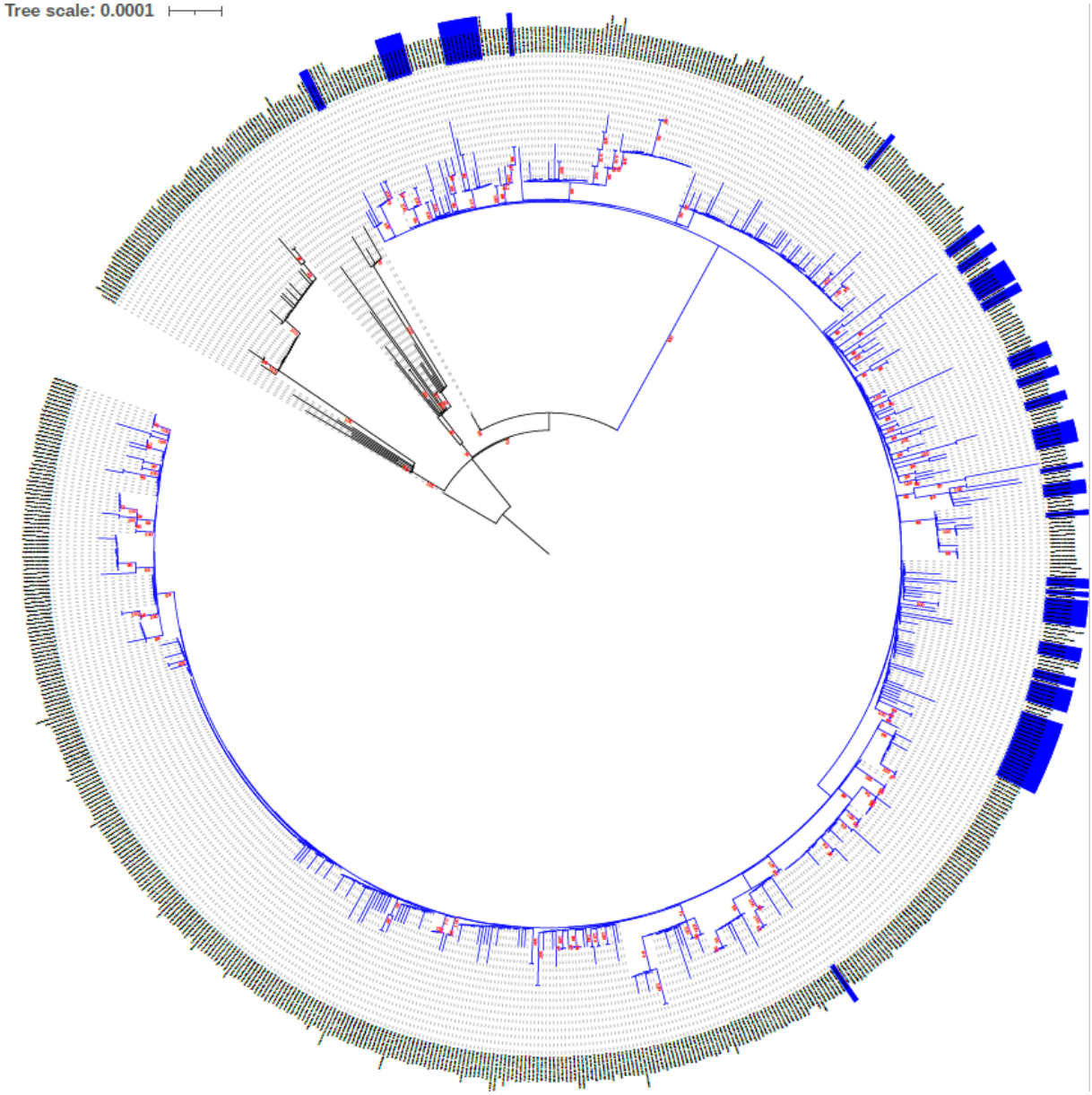
Supplementary Figure S1. Numbers of SARS-CoV-2 Marseille-4 variant genomes deposited in the GISAID database and available in our sequence database according to the country of origin



Supplementary Figure S2. Mutations, amino acid substitutions and diversity in SARS-CoV-2 Marseille-4 strains.



Supplementary Figure S3. Phylogenetic tree based on SARS-CoV-2 full-length genomes. Phylogenetic tree reconstructed from full-length viral genomes obtained from clinical samples. Phylogenetic trees were reconstructed by using the GISAID TreeTool in v2.0 that performs an initial approximate maximum likelihood phylogeny reconstruction using FastTree (Price et al., 2010) then a refinement by RaXML (Stamatakis, 2014).



SUPPLEMENTARY TABLES

Supplementary Table S1. Primers and probe of Marseille-4 variant specific qPCR

Name	Sequence (5'-3')	Positions *
Primers:		
Pri_IHU_ C4_5_MBF	GAGGTTTAGAAGAGCTTTTGGTGA	9,460-9,483
Pri_IHU_ C4_5_MBR	CCAGGTAAGAATGAGTAAACTGGTG	9,549-9,573
Probe (6FAM-labelled):		
Pro_IHU_ C4_5_MBP	CCTTATTTCATTCACTGTACTCTG	9,520-9,543

* in reference to genome GenBank accession number NC_045512.2 (Wuhan-Hu-1 isolate). The nucleotide carrying the mutation specific of the Marseille-4 variant is covered by the probe and underlined.

REFERENCES

Price MN, Dehal PS, Arkin AP**.** FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 2010;5:e9490.

Stamatakis A**.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312-1313.